

GenCore version 5.1.6
Copyright (c) 1993-- 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:04:49 ; Search time 26 Seconds

(without alignments)
3937.814 Million cell updates/sec

Title: AAC39336

Perfect score: 5677
Sequence: 1 MESEGETAGKPMKNIVPQT.....IDPFTSNVTGPDILEGGINC 1065

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5677	100.0	1065	2 T52054	cellulose synthase
2	3984	70.2	1081	2 T05351	cellulose synthase
3	3811	67.1	1026	2 T51579	cellulose synthase
4	3801	67.0	1065	2 F84649	probable cellulose
5	3628	63.9	1084	2 T08583	cellulose synthase
6	3612.5	63.6	1081	2 T52028	cellulose synthase
7	3590	63.2	1088	2 H84604	cellulose synthase
8	3473.5	61.2	974	2 T10797	probable cellulose
9	3317.5	58.4	958	2 T04870	cellulose synthase
10	2808	49.5	685	2 T10800	cellulose synthase
11	2324	40.9	1181	2 D86157	hypothetical prote
12	2303.5	40.6	1145	2 T51546	cellulose synthase
13	2289.5	40.3	1111	2 T05646	hypothetical prote
14	2221.5	39.1	1036	2 D84741	probable cellulose
15	2185.5	38.5	979	2 C86446	cellulose synthase
16	1931	34.0	583	2 T02209	cellulose synthase
17	1092.5	19.2	757	2 T02561	probable cellulose
18	1069	18.8	712	2 T02552	cellulose synthase
19	1043	18.4	755	2 T02553	cellulose synthase
20	1018.5	17.9	748	2 T02360	cellulose synthase
21	953.5	16.8	689	2 T08918	hypothetical prote
22	948.5	16.7	727	2 T08920	hypothetical prote
23	939.5	16.5	686	2 T08919	hypothetical prote
24	934.5	16.5	828	2 E71417	hypothetical prote
25	849	15.0	710	2 B7147	hypothetical prote
26	441.5	7.8	693	2 AF2275	hypothetical prote
27	305.5	5.4	326	2 T08591	hypothetical prote
28	301	5.3	322	2 T12093	hypothetical prote
29	283.5	5.0	322	2 T08592	hypothetical prote

30	273	4.8	874	2 AB0985	probable polysacch
31	273	4.8	888	2 E91180	probable cellulose
32	272	4.8	692	2 S47754	hypothetical prote
33	272	4.8	888	2 H61511	hypothetical 101.6
34	269	4.7	888	2 G68026	probable cellulose
35	265	4.7	759	2 D70422	cellulose synthase
36	252.5	4.4	1596	2 T11338	cellulose synthase
37	245.5	4.3	729	2 A98320	cellulose synthase
38	245.5	4.3	729	2 AD2963	cellulose synthase
39	245.5	4.3	861	2 T19714	cellulose synthase
40	242.5	4.3	664	2 A95889	cellulose synthase
41	228.5	4.0	754	2 A43735	bcsA protein - Ace
42	196	3.5	723	2 C36963	cellulose synthase
43	178	3.1	654	2 D98350	hypothetical prote
44	178	3.1	654	2 AB2932	beta 1,3 glucan sy
45	173	3.0	768	2 B97083	glycosyltransferas

ALIGNMENTS

RESULT 1									
T52054									
cellulose synthase (EC 2.4.1.-) catalytic subunit [validated] - Arabidopsis thaliana									
C:Species: Arabidopsis thaliana (mouse-ear cress)									
C:Date: 20-Oct-2000 #sequence: revision 20-Oct-2000 #text: change 31-Dec-2000									
C:Accession: T52054									
R:Arifoli, T.; Peng, L.; Betzner, A.S.; Burn, J.; Wittke, W.; Herth, W.; Camilleri, C.									
Science 279, 717-720, 1998									
A:Title: Molecular analysis of cellulose biosynthesis in Arabidopsis.									
A:Reference number: 213745; PMID:98111412; PMID:9445479									
A:Accession: T52054									
A>Status: Preliminary; translated from GB/EMBL/DBJ									
A:Molecule type: mRNA									
A:Residues: 1-1065 <ARI>									
A:Cross-references: EMBL:AF027174; PIDN:AAC39336.1									
C:Genetics:									
A>Note: Ath-B									
C:Function:									
A:Description: EC 2.4.1.-; cellulose synthase [validated, PMID:98111412]; involved in									
C:Keywords: glycosyltransferase; hexosyltransferase									
Query Match									
Best local similarity 100.0%; Score 5677; DB 2; Length 1065;									
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MESEGETAGKPMKNIVPQT	ICSDNNGKTVGDRFVACDICSFPVCRPCYEYERKDNQ	60					
DB	1	MESEGETAGKPMKNIVPQT	ICSDNNGKTVGDRFVACDICSFPVCRPCYEYERKDNQ	60					
QY	61	SCPOCKTRYKRLKSPAI	PGDKDEGLADEGTVERNYPQKEIKERMLGWLTRGKGEM	120					
DB	61	SCPOCKTRYKRLKSPAI	PGDKDEGLADEGTVERNYPQKEIKERMLGWLTRGKGEM	120					
QY	121	GEPOYKVEVSHNHL	PRLTSRDSTGSEFSAAPRELVSTTAGKRLPYSDVNOQSPRR	180					
DB	121	GEPOYKVEVSHNHL	PRLTSRDSTGSEFSAAPRELVSTTAGKRLPYSDVNOQSPRR	180					
QY	181	IVDPVGLGNVAMKER	VDGWMKKQKNTGPTQAAERGVGDIDASTDILDAELNDEA	240					
DB	181	IVDPVGLGNVAMKER	VDGWMKKQKNTGPTQAAERGVGDIDASTDILDAELNDEA	240					
QY	241	ROPISRKVSIPSSRI	NPYRWIMRLVILCLFLHYRTNPPVNAFALMLVSVICETIPAL	300					
DB	241	ROPISRKVSIPSSRI	NPYRWIMRLVILCLFLHYRTNPPVNAFALMLVSVICETIPAL	300					
QY	301	SWIIDPQPKMPVRE	TYLDRLARYDREGSPQLAANDIVSVDPDKREPLYTANTVL	360					
DB	301	SWIIDPQPKMPVRE	TYLDRLARYDREGSPQLAANDIVSVDPDKREPLYTANTVL	360					
QY	361	SILAVDPVDKVC	YVDDGAAMLSPESLAETSEFARKWVFCCKYSIEPRAPMYFAAK	420					
DB	361	SILAVDPVDKVC	YVDDGAAMLSPESLAETSEFARKWVFCCKYSIEPRAPMYFAAK	420					

Oy 421 IDYLKDKVQTSFVVDRAAMKREYEFPKRIKRIALVSKLKCEBGMWDDGPMFGNNMGD 480
 Db 421 IDYLKDKVQTSFVVDRAAMKREYEFPKRIKRIALVSKLKCEBGMWDDGPMFGNNMGD 480
 Oy 481 HPGMIQVFLGONGSLDAEGNELPRLVYVSRKRGPFQHHKKAGAMNLVRYSAVLTNGPF 540
 Db 481 HPGMIQVFLGONGSLDAEGNELPRLVYVSRKRGPFQHHKKAGAMNLVRYSAVLTNGPF 540
 Oy 541 ILNLDCOHYINNSKALREAWCFELMDPNLGRKVCYVQPPQRDGDIDKNDRYANRNTVFEDI 600
 Db 541 ILNLDCOHYINNSKALREAWCFELMDPNLGRKVCYVQPPQRDGDIDKNDRYANRNTVFEDI 600
 Oy 601 NLRLGLDGIQGVVYGTGCVFNRTALYGEPRIKHKHKPSSLKSLKCGSRKKNSKAKES 660
 Db 601 NLRLGLDGIQGVVYGTGCVFNRTALYGEPRIKHKHKPSSLKSLKCGSRKKNSKAKES 660
 Oy 661 DKRSGRHTDSTVYFVFNLDIEEGVEGAGFDEKALMSQMSLEKRFQGSAPVASTLME 720
 Db 661 DKRSGRHTDSTVYFVFNLDIEEGVEGAGFDEKALMSQMSLEKRFQGSAPVASTLME 720
 Oy 721 NGVVPSPATPNLKEAIHVISCGYEDKSDMGMEIGMTYGSVTEDDILTFGMHARGRSI 780
 Db 721 NGVVPSPATPNLKEAIHVISCGYEDKSDMGMEIGMTYGSVTEDDILTFGMHARGRSI 780
 Oy 781 YCMKPLPAFKGSADINLSDRLNQLRWALAGSVEILFSRHCPIWYNGRLKFLERFAYVN 840
 Db 781 YCMKPLPAFKGSADINLSDRLNQLRWALAGSVEILFSRHCPIWYNGRLKFLERFAYVN 840
 Oy 841 TTITPISIPILMCTLLAVCLFTNOFTIPQISINASTWPLSTLSTFANGILEMRSGV 900
 Db 841 TTITPISIPILMCTLLAVCLFTNOFTIPQISINASTWPLSTLSTFANGILEMRSGV 900
 Oy 901 GIDEMWRNEQFVWIGVSAHLFAVFOGILKVLADIDNTFTVTSKASDEDDGFALTYLFKW 960
 Db 901 GIDEMWRNEQFVWIGVSAHLFAVFOGILKVLADIDNTFTVTSKASDEDDGFALTYLFKW 960
 Oy 961 TTLILPPTLLIIVNLGVAVGVSTAINSGVSKGRLFFAFAWVYVHLPLKGLMGR 1020
 Db 961 TTLILPPTLLIIVNLGVAVGVSTAINSGVSKGRLFFAFAWVYVHLPLKGLMGR 1020
 Oy 1021 QNRPTTIVVWSVYLASFSLMWIRIDPFTSRVGPDLLEGGINC 1065
 Db 1021 QNRPTTIVVWSVYLASFSLMWIRIDPFTSRVGPDLLEGGINC 1065
 RESULT 2
 105351
 cellulose synthase (EC 2.4.1.-) catalytic chain RSW1 - Arabidopsis thaliana
 N:Alternate names: protein F8B4.110
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
 C:Accession: 105351
 R:Bayan, M.; Terlyn, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De
 submitted to the Protein Sequence Database, February 1999
 A:Reference number: 215409
 A:Accession: 105351
 A:Molecule type: DNA
 A:Residues: 1-1081 <BEV>
 A:Cross-references: EMBL:AL034567
 A:Experimental source: cultivar Columbia; BAC clone F8B4
 C:Genetics:
 A:Gene: RSW1
 A:Map position: 4
 A:Introns: 27/3; 93/1; 150/3; 189/1; 252/2; 341/2; 456/3; 502/3; 544/3; 704/1; 766/1
 A:Note: F8B4.110
 C:Keywords: glycosyltransferase; hexosyltransferase; P-loop
 Query Match 70.28; Score 3984; DB 2; Length 1081;
 Best Local Similarity 69.78; Pred. No. 4.6e-307;
 Matches 741; Conservative 128; Mismatches 158; Indels 36; Gaps 13;

QY	2	ESEETAGKPKRNILVPOQCGSPNNQKGVGDVDFVACDICSFPVRCYCYERKDGQOS	61
Db	23	ESBDGT--KPLAKNNGOICQICSDVDGLAETGVDVFNACNECAFVYCRICYEERKDGQOC	80
QY	62	CPQCKTRKRLKGSIPAIPGDKDEGLADGTEVEFNYDQEKISERMLGMILTRKGEMG	121
Db	81	CPQCKTRFRRRRSGSPRVGEDDEDV--DDEINENYAO-----GANKRRHORHG	128
QY	122	EPOYDKEVSHNH--LPLRTSGQODSGEBSAASPERLSVSSP---IAGKRLPYAS---DY	173
Db	129	E-EFSSSSRHSQDPIPLTLHGHTVSGEL--RTPOQSRITSGPLGPSDRALISSPYIDP	185
QY	174	NQSPNRRIYDP-----VGLGNVAMKERVNDGMMKKQKQKXNTGTPVSTOASERGVYDIDAST	227
Db	186	RQPYRPAVRIYDPSKOLNSYGLGNVDMKERVEBQWKLOKQNMLOMGKXHEGAG-EIE-GT	243
QY	228	DIADLEALINDEANQPLSRKYSIFSSSRINPVRMYIMLRVLICFLYHTRITNPPRNAPAL	287
Db	244	GSNNEELQMDATLPPMSRVVPISSSRILTPPRVYIILRLILCEFLDYRTNPKPNVPL	303
QY	288	WLVSVICEIMFALSMIIDOPFKMPVRNRETYLDRALYHDEGEPSSOLAADIVFVSTVPD	347
Db	304	WLTVSVICEIMFASWMLDOPFKKPIPIRETYLDRALYHDEGEPSSOLVPRDYVEYSTVPD	363
QY	348	LKEPPLVTANTVLISILAVDYPVDKVCYFEDDGAAMLSFESLAEITSEFARKVDPCKKYS	407
Db	364	LKEPPLVTANTVLISILSVDPYVDKVCYACVSDGASAMLTSELSAEFAKKWPFCKKFN	423
QY	408	IEPAPRPMYPAKIDYLLKDKVQTSFVNDRAKMKREYEERFKRIANALVSKLCPEEGVVM	467
Db	424	IEPAPRPEYFQAKIDYLLKDKIQPSFVYERRMKKREYEERFKRIANALVAKAKITDEBGTM	483
QY	468	QDGTWPWGNNNGDPRGMIOVFLGONGGLADABAGNLPRUVVYSREKRPQFQHNKAKAGANA	527
Db	484	QDGTWPWGNNTRDHPGMIOVFLGSHSGLDITDGNELPRLIYYSREKRPQFQHNKAKAGANA	543
QY	528	LVRISAVLTNPFLTLNDCHYIINNSKALBRAMCFLMDPNLNGKOVCYQVOPRRDGDIGN	587
Db	544	LIRISAVLTNCAVLTNDCDHYFNNSKAIKAMCFMDPRALGKKCCVYQFOPRRDGDIDLH	603
QY	588	DRYANRMTVPFDILNRGLDIOGVYUVTGCVFNRTALYGEPRILYKHNKRPSSLKSLCG	647
Db	604	DRYANRMTVPFDIMMKGIDIOGVYUVTGCVFNROALYGDVPLTEDEDLNIIYVSSC	663
QY	648	GSRRKNSKAKKESQKKKSG-RHDSYTPVFNLDIEBGVAGFDEDEKALLMSQMSLEKR	706
Db	664	GSRRKNSKKSXKYNKEGRKIRNSQSNAPLFNMEIDEGFE--GYDDESIILMSQSVYERK	721
QY	707	FGQSAVAVPASTLMMGGVGPSPAPTEBNLLKAIHYSISGEYDKSOMQMEIGMTYSVEDI	766
Db	722	FGQSPVPEIATFEMQGGIPPTTNPATLKEAIHYSISGEYDKTEKKEIGMTIYSSVEDI	781
QY	767	LTFGRKMARGRNSIYCMKLPAPFKGSAPINLSDBLNOVLRWALGSVELTFSRHCPIMWGY	826
Db	782	LTFGRKMARGRNISIYCNBPAPFAFGSAPINLSDBLNOVLRWALGSIELTFSRHCPIMWGY	841
QY	827	NGRLKFLERPAVYVTTIYPTISIFPLMTCITLAVCLPTNQPITPOISINASTWLSLFLS	886
Db	842	HGRILTERLRIAYIMTYIPTSISPLIAYCIIIPACCLITDRPIEISINASTWLSIWTLLTIS	901
QY	887	IFATGILEMRHSGVIDEMWNRDEQFWYIGVSAHLFVFOCILLVLAGIDINFPNTSKAS	946
Db	902	IATVIGILELRKSGVISIEDMMNRDEQFWYIGISAHLPFAVQGLLAVLAGIDINFPNTSKAT	961
QY	947	DEDDGFAELYLKFWTTLIPETTLIYNLVGVVAGVSAIINSQYQSMGPLGKLEFAFW	1006
Db	962	DEDDGFAELXYLFKWTALLIPETTVLLNVLIGIVAGVSAVNSGYQSMGPLGKLEFALMV	1021
QY	1007	IVHLYPFLKGLMGQONRPTIIVYVSVLASISFSLWVRIDPF	1049
Db	1022	IAHLYPFLKGLMGQONRPTIIVYVSVLASISFSLWVRIDPF	1064


```

0Y 339 JLEFVSTVPLKEPPLVYANTVSTIIAYDVPVKVSTYVEDDDGAAMLSFSLSIAETSEFARK 398
Db 342 DVFSVTPDMKEPPLVYANTVSTIIAYDVPVKVSTYVEDDDGAAMLFEHLSIAETSEFARK 401
0Y 399 WVEPCKKYSIEBRAEWMYFAAKIDYLDKQVQTSFVMDRRAMKREYEEFTRIAYLSKAL 458
Db 402 WVEPCKKFNIEBRAEWFYFSQKIDYLDKQISFSFVERAMKREYEEFKRINIYLAKAO 461
0Y 459 KCPREGWMOOSTPMPDGNNGDHPGMOIOVEFLONGSLDMDGNLPLVYVSRKRGQFH 518
Db 462 KIPDEGWMDEBTSMPGNAPRDPHGPQIOVEFLHSGSLDMDGNLPLVYVSRKRGQFH 521
0Y 519 HKKAGAMALVAVSVLITNGPPLIIMDCCHYIINNSKALREAMFLMDPNLGKQCVYQF 578
Db 522 HKKAGAMALILVSAVLITNGAVLLANVDCOHYFNNSKALREAMCFMMDPAIGKKCCVYQF 561
0Y 579 QRFQSDIDKNDKRAANNVYFEDINLGLDIOGPVYVGTGCVNRTALYGEPIYKXHKK 638
Db 562 QRFQSDIDLADRANNTVYFEDINLGLDIOGPVYVGTGCCNRAOLYGDVPLTEEDLE 641
0Y 639 PSLSLKLCGGSKRKKRNSKAK--KESDKKKSGRHNDSTVPVFNLDIEEGVGEAGFDDSKALL 697
Db 642 PNIIYSCQFSGSKKKSKRKIRIYEDNRSLIKRSDSNVPLENMEIDEDVE--GEYDEMSLL 699
0Y 698 MSQMLEKRFQGSANFVASTLMENGSVPSSAIPENILKALIHVISCYGEDSKSPWMEIGW 757
Db 700 VSQKLEKRFQGSPEFIATFMEQGLPSTPNLPLILKALIHVISCYGEAKTDMGEIGW 759
0Y 758 IYGSVTEDLITFCFKHARGMRSIYCPKLPFAFGSAPINLSDRLNOVLKALGSVELTFS 817
Db 760 IYGSVTEDLITFCFKHARGMRSIYCPSPARKGSAPINLSDRLNOVLKALGSIELLS 819
0Y 818 RHCPITWYGNGLKFLERFAYVNTTYITPSIPLMYCPLLAVCLEFTNOFIIPQISNLAS 877
Db 820 RHCPITWYGNGLKFLERFIAYINTFYITPSIPLAYCMLPAFCLITNFIETISNLAS 879
0Y 878 IWFLSLPISTIPATGILEMWSGVGIDEMWRNDFOWYIGVSAHLFAPQGLIKVLAGIDT 937
Db 880 LCFMLLFASITASAILLEKMSDYALFEDMWRNEFPWYIGTSALHFAVQGLLEVFGIDT 939
0Y 938 NFFVYISKASDEDEGDAELIYFKWTLILPPTLLIYNLGVVAGVSYAINSQSGWGPLF 997
Db 940 NFFVYISKASDEDEGDAELIYFKWTSLLIPPTLLIYNLGVVAGVSYAINSQSGWGPLM 999
0Y 998 GKLPFAFVYIYVLVYFLKGMGRORPTIYVYVWSVLASISLSLWVRIDPFSIRYTG 1055
Db 1000 GKLPFAFVYIYVLVYFLKGLGRONTPTIYVWSVLASISLSLWVRINPFS--FTG 1056

```

RESULT 5

cellulose synthase (EC 2.4.1.-) catalytic chain - Arabidopsis thaliana
N:Alternate names: protein T22F8.250
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C:Accession: T08583; T09014
R:Byevan, M.; Zimmerman, W.; Gruenisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; May
submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16442
A:Accession: T08583
A:Molecule type: DNA
A:Residues: 1-1084 <BEV>
A:Cross-references: EMBL:AL050351; GSPDB:GN00062; ATSP:T22F8.250
A:Experimental source: cultivar Columbia; BAC clone T22F8
R:Arrioli, T.; Peng, L.; Betner, A.S.; Bun, J.; Witke, W.; Hersh, W.; Camilleri, C.; H
Science 279, 717-720, 1998
A:Title: Molecular analysis of cellulose biosynthesis in Arabidopsis.
A:Reference number: Z13745; MUID:98111412; PMID:9445479
A:Accession: T09014
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1084 <ARI>

A:cross-references: EMBL:AF027173; NID:g9827140.; PID:n92837141
A:Experimental source: cultivar Columbia
C:Genetics:
A:Gene: ATSP.T22F8.250; Ath-A
A:Map position: 4
A:introns: 2/3; 156/3; 191/1; 254/2; 343/2; 458/3; 504/3; 546/3; 617/3; 708/1; 773/3
C:Keywords: cell wall synthesis; glycosyltransferase; hexosyltransferase

Query Match	63.9%	Score 3628	DB 2	Length 1084
Best Local Similarity	63.0%	Pred. No. 8e-279		
Matches 678	Conservative 165	Mismatches 191	Indels 42	Gaps 16

OY	10	KPKNNIPQOCICSONVKTQYDGRFPAFCIDICSPCPYCPCEYERKDQNOQOCQSTRY	69
Dd	29	RSVOELSGOTCQICGDEILITVSSLEFPAACNECAPVPCPYEYERREBGNACQOCSTRY	88
OY	70	KRLGSPALPGDKDEGLADEGTEVEFN - POKEKISERMLGMHLTRKGEMGEPOYDKE	128
Dd	89	KRIKSGPRVDGDEDEEDIDLEYEFDHGMDEPHAAEALSSRLNTGNG-----GLDSA	142
OY	129	VSHNHLPRLTSGRODTSGEFSAASPERLVSSTIAGAKL---PSSDYNOQSPNRIYDP-	184
Dd	143	PQSQJPLPLTYCEDEDMTS--DRHALYPPSTGNGNVYPAPE-TDSSAPPAQSAWVPO	199
OY	185	-----VGLGVAMKEKEDGMMKOKKNGNPVSTOAAE-RCGVNIDASTOLDLEADLND	238
Dd	200	KDIAEYGVSAVAKMDMEYWKRRQGEKQVYIKHEGNGNGRSNDDELD---PDMAMD	256
OY	239	ENAPOLSRKKSIPSSRINPYRWIMRLVLIICLFPHYRTIPVNAEALMVSYICEIWF	298
Dd	257	EGNOPLSRKPIRSSHINPYRMLIICRLALIGLFPHYILHPVADATGLMILTVCIGIWF	316
OY	299	ALSMLIDQEFKMEPVNRETYDLRLALRDGSEPSQLAANDIESTYVDPLKEPPLVTANT	358
Dd	317	AVSMIDQEFKWPPIRETYDLRLSLREYKEKSPGLGADVDFVSTVDPLKEPPLITANT	376
OY	359	VLSILAVDPVOKSVCTYFDDCAAMLSFSLAEFSEFARKVNPCKKYSTIEPRAPDEYFA	418
Dd	377	VLSILAVDPVOKVACYSDDGAAMLTTEALSUDTAEFRARKVNPCKKFNLIEPRAPDEYFS	436
OY	419	AKIDYLKDKQVTSFYVDRBRAMKEVEEFKIRINALVSALKCPREGWMDGTFWPNNT	478
Dd	437	QKMDYLKNKHYPFAVERARMAKRDYDEERVKINALVALAKQVPEEGTMDGTFWPNNV	496
OY	479	GDBRGMIOYFLGONGILDAEGNELPRLYVVSSEKRPGFQHHKKAGAMNALVRSAYLTNG	538
Dd	497	RDBRGMIOYFLGSHGVRODTGNEPLRPLYVVSSEKRPGEFDHKKAGAMNSILRVSAYLSNA	556
OY	539	PFLINDCDHYINNSKALBEAMCFMLDPRMLQOVCTYVOPORFQIDKNDRYANRNTYEF	598
Dd	557	PFLYNDCCDHYINNSKALRESMCFPMNDPQSKKVCYVOPORFQIDIRHRYSNRNVFF	616
OY	599	DINLRGIDLOGVYVYGTCVFNPRALGVBERPIKVRKKP-----SULSK---LGGSR	650
Dd	617	DINNRGIDLOGVITVGTGCVFRROALGFEDP---KKKRPQGTCCMWRKMCCLCGGLR	673
OY	651	KKN-SKAKESDKKKSGRHTDSTVEVFNLDIEEGVEGAGFDDEKAKALLMSOMLSLEKRFQ	709
Dd	674	KKSKTKAK--DKTKTKETSQIHA--LEANDGVIVPVSINVEKRSBEATQILKEKKFGQ	728
OY	710	SAVFVASTMEBNGVPPSATPENLLKEALIHVISCYEDKSDMGMEICMITYGATEDILTG	769
Dd	729	SPFVFAVAVLQNGVPRNAPSACLLREAIQVYISCGYEDKTEMGKEIMIGYSTEDILTG	788
OY	770	EKNHAGMSTYCMPLPAFKGSAPINLSDRINQVLRALGSGVILFSHRCPIMYGNGR	829
Dd	789	EKNHCGMSTYCMPPRAAFKGSAPINLSDRIOVLRALGASVIEPLSRHCPIWYGGGG	848
OY	830	LKLEFEFAVANTYITISPIPLMCTCLLACFLNORLIPOIENIASIFLSLTSIFA	889
Dd	849	LKMLDERSTYNSVYWTSLPIIYVCSLPACVCLLTGRTIYEISINAGILEPMLMFISIAV	908
OY	890	TGLELRWQSGVIGDEWNRDQFWJIGVSAHLFAVFOGILKVLAGIDTNEFTVYSKASDED	949

Db 181 PQRFEDGIDRHDRYANRNVFEDINMLGLDGLGPPYVGTGCVENRQALYGDPPVSEKR 240
 QY 637 KR-----PSILSKLGGSRKKRSKRAKE-----SDKKSGRH--TDSTVPFNL 678
 Db 241 PKMTCCWCCWCCGCCGSRKSKKKEKGLLGLLGGKMKMKNNYKKSASAVFDL 300
 QY 679 DDIEGVEGAGFDD--EKALLMSQMSLEKRGOSAVFVASTLMENGVPSPAPENILKEA 737
 Db 301 EEIEEGLE--GYEELKSTLMSQKNEKRRGQSPVFIASITLMENGGLPEGTNSTLKEA 358
 QY 738 IHVSGYEDKSMGMEIGMIVGYVEDLITGFKMARGRMSYICMPKPAFGSAPINL 797
 Db 359 IHVSGYEKTEKMEIGMIVGYVEDLITGFKMKRGKMSYICVPKPAFGSAPINL 418
 QY 798 SDRLNGLVRLMAGSVILESRHCPIMYNGRLEKFERAYVNTYITPISILNACTL 857
 Db 419 SDRHLVRLMAGSVILESRHCPIMYNGRLEKFERAYVNTYITPISILNACTL 478
 QY 858 LAVCFEFTNPIIPQISINIASIMFLSLPISIFATGILEMRMSGVGIDEMMRNEQFWVIGV 917
 Db 479 PAVCLLTGKFIITPTLSNLSVWFALFLSIATGVELMRMSGVSIDMMNRNEQFWVIGV 538
 QY 918 SAHLEFVPGIILKVLADIDNFTYNSKASDEGDAFELYFKTTLITPTLLIYNLNG 977
 Db 539 SAHLEFVPGIILKVLADIDNFTYNSKASDEGDAFELYFKTTLITPTLLIYNLNG 597
 QY 978 VVAGVSYAINSGYOSGMPPLSKLFFAFWVIVHLVPLKGLMGRONTPRTIVVMSVLAS 1037
 Db 598 VVAGVSDAINNCGSMGSPPLSKLFFAFWVIVHLVPLKGLMGRONTPRTIVVMSVLAS 657
 QY 1038 IESLWVRIDPPTSRVTGPDILEGINC 1065
 Db 658 IESLWVRIDPPTSRVTGPDILEGINC 685

RESULT 11

Db6157
 hypothetical protein F22D16.26 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
 C:Accession: D86157
 R:Theologos, A.; Ecker, J. R.; Palm, C. J.; Federspiel, N. A.; Kaul, S.; White, O.; Alonso, C.; Chiu, C. M.; Chung, M. K.; Conn, L.; Conway, A. B.; Conway, A. R.; Creasy, T. H.; Dewar, K.; Hansen, N. F.; Hughes, B.; Hultzer, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J. L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C. A.; Li, J. H.; Li, Y.; Lin, S. X.; Liu, Z. A.; Luros, J. S.; Maitl, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S. L.; Schwartz, J. R.; Shinn, P.; Southwick, A. M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C. M.; Venter, J. C.; Davis, R. W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: D86157
 A:status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1181 <STO>
 A:Cross-references: GB:AE005172; NID:96056428; PIDN:AAF02892.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 40.98; Score 2324; DB 2; Length 1181;
 Best Local Similarity 44.0%; Pred. No. 2,4e-175;
 Matches 491; Conservative 165; Mismatches 268; Indels 192; Gaps 29;
 QY 18 QTCQI--CSDNGKTYDGRFVADICSPVCRPCYEYERKDNQSCPOCKTKYKRLKS 75
 Db 160 QTCMLGCGDE--KVVHG---RCE--CGFRICRDCYFDCITSGGGCPCGCKEYRIND 211
 QY 76 PAIPGDDDEGLADESTVEFNYPQKEKISERMLGMHLTRGKGEMGEPOYDKESHNL 135
 Db 212 PEEDEDEDEDEKPP-----LPQ-----MSESKIDKRLS---VY 241
 QY 136 RLTSRODTSGEFSASPERLSVSTIAGKRLPLSSDVNOSPNRRITVDPGLGNVAMKER 195

Db 242 KSEKAOQAGDEPDHT-----RMLEFK-----GTGYGNAMVPR 275
 QY 196 VQGMKKQKQKNGPVSTQASERGVYDIDASTDILADEALLNDEAROPRLSPSSRI 255
 Db 276 -DGYI-----GSGGGNGYEPPE-----FGRSRKRLTKVYSAAII 314
 QY 256 NPYRATIMRLVILCLFLHYRTNPVPAFALMVSVICEIMFALSWIIDOPKPPVNR 315
 Db 315 SPYRLIALRLVALGLFLWRVRHPRRLEAMMLMGSTCELPWFALSWIIDOPKPPVNR 374
 QY 316 EYLDRLALRYDEG--EP---SOLAVIDFVSTVDPLKEPLVYANTVLSIANDY 370
 Db 375 LFDLGLVKERFESPNLRNPKGRSDLPGLVFPVSTADPEKPEPLVYANTVLSIANDY 434
 QY 371 KVCYVEDDGAALMSPESLAETSEFARKVPPCKKTSIRPEWFAKIDYLRKYOT 430
 Db 435 KLACYSDDGALLTEALAOIASFASVWPCCRKNIIEPRNDEATFGOKRNLKAVRL 494
 QY 431 SPYKDRAMKREYEERKINML-----VSKALC----- 460
 Db 495 DFYERRRRYKREYDEKVRINSLPEAIRRSDAVYVHEELRAKKQKQKMMGNPQETVI 554
 QY 461 -PEEGVMODGTPWPG-----NTGCHPMIOVFLQNG-----GLDAEGN----- 500
 Db 555 VPKATV-MSDGSHPGPTWSSGETNSRGDAGIIOQLMPPNAEPYGAADAENLIDPT 613
 QY 501 ----ELPRIVYYSREKRGFOHKKKAGANALVRSASVLTNGPFILNLCDDHYINNSKAL 556
 Db 614 DVDIRLPMALVYSREKRGFOHKKKAGANALVRSASVLTNGPFILNLCDDHYINNSKAL 673
 QY 557 REAMCEFLMDNLGQYVQVQFQRFQDIDKNDRYANRNVFEDINMLGLDGLGPPYVGT 616
 Db 674 RBGCMFLDRG--GDRICYVQFPQRFEGIDPNDRYANHNTVFPVYSRALDGLGPPYVGT 732
 QY 617 GCYFNRTALXGYEPPKVKHKKPSLSKLCCGSRKNSAKKESDKKSGRTDVTVP 676
 Db 733 GCIFRTALXGSPPRATHH-----GWLGRKRVVILSRPRAMKKKDEVSLEPLN 783
 QY 677 NLDIDIEGVEGAFDEKALMSQMSLEKRGOSAVFVASTLMENG-- 722
 Db 784 GEYNEENDDG-----DIESL-----LPRFGNSNFVASIPVAYEGRLIDDLQKGN 834
 QY 723 -----GVPSATPENLKEALHIVISCYEDKSDMGMEIGMIVGYSTEDILTFEKMAR 775
 Db 835 SRPAGSLAVRPELDATVAEALSVISCYEDKTEKRGVIGSVTDDVYGYRMNR 894
 QY 776 GMRSTYCMKRLAFKCSAPINLSDRLNOLVRLMAGSVLELFSRHCIYWGYNRLKFLER 835
 Db 895 GMRSTYCMKRLAFKCSAPINLSDRLNOLVRLMAGSVLELFSRHCIYWGYNRLKFLER 952
 QY 836 FAVNNTYIPITSIPPLMCTLLAVCLFTNPFPIIPISINIASIMFLSLFISFATGILM 895
 Db 953 VAFVNGMTPFSLFLIYVCIPLAISLFGQFVOSLDTLLIYLISLITLCLMSLLEI 1012
 QY 896 RMSGVGIDEMMRNEQFWVIGVSAHLFAVFGILKVLADIDNFTYNSKAS--DEGDD--F 952
 Db 1013 KMSGITLHEMMNRNEQFWVIGVSAHLFAVFGILKVLADIDNFTYNSKAS--DEGDD--F 1072
 QY 953 AELYLEFKWTTLITPTLLIYNLGVVAGVSTAINSGYOSGMPLEKLFPAFVIVHLXP 1012
 Db 1073 ADLYVVKMSFLWVPLTIMVNMVIALVAGIARLVSPFOWMSKLVGVGFSPWVLCILXP 1132
 QY 1013 FLKGLMGRONTPRTIVVMSVLASLFLSLWVRIDP 1048
 Db 1133 FAKGLMGRONTPRTIVVMSVLASLFLSLWVRIDP 1168

RESULT 12

T51546
 cellulose synthase catalytic subunit-like protein - Arabidopsis thaliana
 N:Alternate names: protein F2K13_60
 C:Species: Arabidopsis thaliana (mouse-ear cress)

[illegible]

796 VKNGRPPALLIIPRELLDASTVAEAIATVISCATEDKTEGSRIGMTYIGSTEDVGYCRM 855
 QY 773 HARGMSRIYCEPKLPEAFKGSAPINLSDRLNOVLRMALGSEVLEFSRRCPIWGYNGRLK 832
 Db 856 HNRGKSVSYCYTKRPAFGSTAPINLDRHQVLRLMATSVEIFEFSNNMLL--ASSKMKI 913
 QY 833 LERRAYVNTTYPITSIDPLMCTLLANCLFNNQFIIPDISNATSWFSLFSLTATGTI 892
 Db 914 LQRIAYLVNGVYIPFETISFLIYCFPLASLFEQGFQVQLNWFLEVLIIISTLCIAL 973
 QY 893 LEMRSGCIDEMWNEQFVWIGVSAHLFAVFOGLIKYLAGIDTFWTYSKA--SDEG 950
 Db 974 LEIWSGISLEEMWNEQFVWIGVSAHLFAVFOGLIKYLAGIDTFWTYSKA--SDEG 1033
 QY 951 DPAELYLEKMTTLLIPTTLLILVNLGVYAVGSAVYAINSGVSGPLFGKLPFAFVAVHL 1010
 Db 1034 EFADLYWKMYSIMIRPTTIIMVNLAIANGVSRITISVVPQMSKILGIVFSEFWLAHL 1093
 QY 1011 YPELGLMGRONRPTIYVWMSVLLASIFSLMVRIDP 1048
 Db 1094 YPFAKGLMGRGRPTIYVWMSGLVATISLMAVAINP 1131
 RESULT 13
 T05646
 hypothetical protein F20D10.310 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
 C:Accession: T05646
 R:Byevan, M.; Medler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer
 submitted to the Protein Sequence Database, February 1999
 A:Reference number: Z15420
 A:Accession: T05646
 A:Molecule type: DNA
 A:Residues: 11111 <BEV>
 A:Cross-references: EMBL:AL035538
 A:Experimental source: cultivar Columbia; BAC clone F20D10
 C:Genetics:
 A:Map position: 4
 A:Introns: 139/22: 675/3
 A:Note: F20D10.310
 Query Match 40.3%; Score 2289.5; DB 2; Length 1111;
 Best Local Similarity 43.1%; Pred. No. 1.2e-172;
 Matches 481; Conservative 158; Mismatches 256; Indels 221; Gaps 28;
 QY 23 CSDNVGKTVDDRFVACDTCSPVCRPCYEYERKDNOSCPQCKTKRKLRKSGPAIPDGK 82
 Db 120 CDGVNMKDERGKDVMPCE-CRFKICRDCFMDAOKE-TGLCPGCKEQYK-----IGDL 169
 QY 83 DEDGL-ADEGTVVEFNPQKEKISERMILGWHLRGKEENGEPQYDKVEVSHNHLPRLSRQ 141
 Db 170 DDDPDPDYSSGGLPLPAFGKDO-----RGNNNNS-----MKR 202
 QY 142 DTSGEFSASPERLSVSSSTIAGKRLPYSSDVNQSERRIVD--DVGLGNVAM-KERYD 197
 Db 203 NQNGEF-----DNRMLFEQGTGYGVGNATWMPQDEM 234
 QY 198 GWAKKQEKNTGPGVSTQAASERGGVIDASTDLADEALLDEAKRPLSRKVSIPSSRINP 257
 Db 235 GDDMDEGRGMVETA-----DKPWRPLSRIRIPPAIISP 270
 QY 258 YRMVIMLRVLVLCJLHYRITNPVNAPALMLVSVICETIMPLALSWILDPKPKFNVNRET 317
 Db 271 YRLIIVIFEVVLCFFLWLRKIRNPEDAIWMLMSTICELMFGESWILDDQPKLCPINRST 330
 QY 318 YLDRLALRYDEGEF-----SOLAANDIFVSTVDLPKEPPLVANTVLSILADVDPVK 372
 Db 331 DLEVLRDKFDWMPSPNPGRSDLPGLDIFVSTADPEKEPPLVANTVLSILADVDPVK 390
 QY 373 SCYVFDGAMLSFSLAETSFAKRWVPCKKYSIEPPAPEMYFAAKIDYLRKQYQTSF 432
 Db 432 SCYVFDGAMLSFSLAETSFAKRWVPCKKYSIEPPAPEMYFAAKIDYLRKQYQTSF 432

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:08:04 ; Search time 32 Seconds

(without alignments)
3558.288 Million cell updates/sec

Title: AAC393336

Perfect score: 5677

Sequence: 1 MESEGETAGKPKMKNIPTQC.....IDPFTSRVATGPDILEGIC 1065

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5677	100.0	1065	US-09-900-237-33	Sequence 33, Appl
2	4507	79.4	1080	US-09-900-237-30	Sequence 30, Appl
3	4133.5	72.8	881	US-09-838-539-8	Sequence 26, Appl
4	3999.5	70.5	1091	US-09-900-237-26	Sequence 26, Appl
5	3922	69.1	1148	US-09-900-237-4	Sequence 4, Appl
6	3873.5	68.3	1086	US-09-900-237-18	Sequence 18, Appl
7	3843.5	67.7	1165	US-09-900-237-8	Sequence 8, Appl
8	3804.5	67.0	1039	US-09-900-237-16	Sequence 16, Appl
9	3483.5	61.4	701	US-09-900-237-32	Sequence 32, Appl
10	3473.5	61.2	974	US-09-838-539-6	Sequence 6, Appl
11	3216	56.6	793	US-09-900-237-18	Sequence 18, Appl
12	3003.5	52.9	740	US-09-900-237-24	Sequence 24, Appl
13	2808	49.5	685	US-09-900-237-31	Sequence 31, Appl
14	2803	49.4	685	US-09-838-539-7	Sequence 7, Appl
15	2647.5	46.6	610	US-09-900-237-16	Sequence 16, Appl
16	2420	42.6	506	US-09-900-237-20	Sequence 20, Appl
17	2055.5	36.2	431	US-09-900-237-28	Sequence 28, Appl
18	1800	31.7	720	US-10-260-046-18	Sequence 18, Appl
19	1405	24.7	522	US-10-260-046-2	Sequence 2, Appl

20	1405	24.7	522	9	US-10-260-046-4	Sequence 4, Appl
21	1388.5	24.5	320	10	US-09-900-237-6	Sequence 6, Appl
22	1358.5	23.9	304	10	US-09-900-237-2	Sequence 2, Appl
23	1152.5	20.3	340	10	US-09-900-237-22	Sequence 22, Appl
24	1071	18.9	590	9	US-10-260-046-30	Sequence 30, Appl
25	1044	18.4	727	9	US-10-260-046-24	Sequence 24, Appl
26	1015	17.9	741	9	US-10-260-046-28	Sequence 28, Appl
27	594	10.5	341	10	US-09-900-237-12	Sequence 12, Appl
28	505.5	8.9	166	10	US-09-734-566-154	Sequence 154, App
29	499.5	8.8	159	10	US-09-734-566-40	Sequence 40, Appl
30	498.5	8.8	165	10	US-09-734-566-38	Sequence 38, Appl
31	289	5.1	115	10	US-09-734-566-42	Sequence 42, Appl
32	278	4.9	693	10	US-09-838-539-11	Sequence 11, Appl
33	245.5	4.3	861	10	US-09-838-539-12	Sequence 12, Appl
34	236.5	4.2	756	10	US-09-838-539-10	Sequence 10, Appl
35	196	3.5	723	10	US-09-838-539-9	Sequence 9, Appl
36	118.5	2.1	1177	9	US-10-024-623-20	Sequence 2, Appl
37	118.5	2.1	1177	9	US-09-795-927-2	Sequence 20, Appl
38	115	2.0	357	10	US-09-874-133-27	Sequence 27, Appl
39	107.5	1.9	1221	9	US-10-270-333-60	Sequence 60, Appl
40	107.5	1.9	1325	9	US-10-154-452-6	Sequence 6, Appl
41	107.5	1.9	1437	9	US-10-154-452-2	Sequence 2, Appl
42	107.5	1.9	1437	9	US-09-934-421A-6	Sequence 6, Appl
43	107.5	1.9	1437	9	US-10-162-012-38	Sequence 38, Appl
44	106	1.9	357	10	US-09-874-133-26	Sequence 26, Appl
45	104.5	1.8	910	9	US-10-012-140-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1									
US-09-900-237-33									
Sequence 33, Application US/09900237									
Patent No. US20020120124A1									
GENERAL INFORMATION:									
APPLICANT: Allen, Stephen									
TITLE OF INVENTION: Plant Cellulose Synthases									
FILE REFERENCE: B01170 US CIP									
CURRENT APPLICATION NUMBER: US/09/900,237									
CURRENT FILING DATE: 2001-07-06									
PRIOR APPLICATION NUMBER: 60/092,844									
PRIOR FILING DATE: 1998-07-14									
PRIOR APPLICATION NUMBER: PCT/US99/15871									
PRIOR FILING DATE: 1999-07-13									
PRIOR APPLICATION NUMBER: 09/720383									
PRIOR FILING DATE: 2000-12-21									
NUMBER OF SEQ ID NOS: 33									
SOFTWARE: Microsoft Office 97									
SEQ ID NO 33									
LENGTH: 1065									
TYPE: PRT									
ORGANISM: Arabidopsis thaliana									
US-09-900-237-33									
Query Match									
Best Local Similarity 100.0%; Score 5677; DB 10; Length 1065;									
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
OY	1	MESEGETAGKPKMKNIPTQCISDNVGTVDREVFACDICSFPVCRPCYEYERKDGN	60						
DB	1	MESEGETAGKPKMKNIPTQCISDNVGTVDREVFACDICSFPVCRPCYEYERKDGN	60						
OY	61	SCPOCKTRKRLKSGAATGDKDEGLABEGTYEFNYPQEKISEHMLMHLTRGGEEM	120						
DB	61	SCPOCKTRKRLKSGAATGDKDEGLABEGTYEFNYPQEKISEHMLMHLTRGGEEM	120						
OY	121	GEPOYKEVSHNLPPLTSGROPTSGEFSASPERLSVSSITAGGKRLPTSSDVNOSPNNR	180						
DB	121	GEPOYKEVSHNLPPLTSGROPTSGEFSASPERLSVSSITAGGKRLPTSSDVNOSPNNR	180						
OY	181	IVDPVGLGVANKERVYDGKMKROEKNVTGVSQAASERGVNDIDASTDILADEALINDEA	240						
DB	181	IVDPVGLGVANKERVYDGKMKROEKNVTGVSQAASERGVNDIDASTDILADEALINDEA	240						

Db 181 IYDPEVLGNVAMKERVDMKMKOEKNTGPVSTQASERGVSDIDASTDILADEALLNDEA 240
QY 241 RQPSLRKVSIPSSSRINPYRMVIMRLVILCLFLHYRITNPVPAFALMLVSVICETIMFAL 300
Db 241 RQPSLRKVSIPSSSRINPYRMVIMRLVILCLFLHYRITNPVPAFALMLVSVICETIMFAL 300
QY 301 SWLLDOPPKWFPVNRRETYLDRLALRYDREGESQOLAVIDIEVSTVDPLKEPPLVTANTVL 360
Db 301 SWLLDOPPKWFPVNRRETYLDRLALRYDREGESQOLAVIDIEVSTVDPLKEPPLVTANTVL 360
QY 361 SIILAVYPPVDKVCYFDDGAAMLSPESLAETSEFARKWPPCKKXSIIEPRAPEMVFAAK 420
Db 361 SIILAVYPPVDKVCYFDDGAAMLSPESLAETSEFARKWPPCKKXSIIEPRAPEMVFAAK 420
QY 421 IDYLKDVOTSFYKDRBRAMKREYEERKIRINALVSKALCPREGVWVMDGTPMPGNNTGD 480
Db 421 IDYLKDVOTSFYKDRBRAMKREYEERKIRINALVSKALCPREGVWVMDGTPMPGNNTGD 480
QY 481 HPGMIOVFILGQNGDLAEGNELPRLVYVSREKRPFGQHHKAGAMNALVRSVAVLTNGPF 540
Db 481 HPGMIOVFILGQNGDLAEGNELPRLVYVSREKRPFGQHHKAGAMNALVRSVAVLTNGPF 540
QY 541 IILNDDCHYIINNSKALREAMCFIADBNILGKVQCYVOFPORFDGIDKNDRYANNTVFEED 600
Db 541 IILNDDCHYIINNSKALREAMCFIADBNILGKVQCYVOFPORFDGIDKNDRYANNTVFEED 600
QY 601 NLRLGDLGIGPVYVGTGVFNRTALYGBPPIKVKKHKKPSLSKLKSGSRKKSKAKKES 660
Db 601 NLRLGDLGIGPVYVGTGVFNRTALYGBPPIKVKKHKKPSLSKLKSGSRKKSKAKKES 660
QY 661 DKRSRHRDSTVPVNLDDIEEGVAGAGFDDKALLMSQSLKRRGOSAVVASTIME 720
Db 661 DKRSRHRDSTVPVNLDDIEEGVAGAGFDDKALLMSQSLKRRGOSAVVASTIME 720
QY 721 NGVPSAPENILKKAIVHISGVEDKSPMGMEIGMIVGVSVEEDILITGKKMARGRST 780
Db 721 NGVPSAPENILKKAIVHISGVEDKSPMGMEIGMIVGVSVEEDILITGKKMARGRST 780
QY 781 YCPMLPAFKGAPINLSRLNQLVRLMAGSVEILFSRHCPIMVGYNGRLKFLERRAYVN 840
Db 781 YCPMLPAFKGAPINLSRLNQLVRLMAGSVEILFSRHCPIMVGYNGRLKFLERRAYVN 840
QY 841 TTYPTTSTPLMAYCTLLAVCLFTNOFIIPQISINIASIWFLSLSTFATGILEMRSGV 900
Db 841 TTYPTTSTPLMAYCTLLAVCLFTNOFIIPQISINIASIWFLSLSTFATGILEMRSGV 900
QY 901 GIDEMRNEDQFVYIGVSASHLFAVFOGILKVLGIDTNEFTVTSKASDEDDFALYLTKM 960
Db 901 GIDEMRNEDQFVYIGVSASHLFAVFOGILKVLGIDTNEFTVTSKASDEDDFALYLTKM 960
QY 961 TTYLIPPTLLIVNLGVVAVGYSAINSGYSQWGPGLKLFPAFWYIVHLYPELKGIMGR 1020
Db 961 TTYLIPPTLLIVNLGVVAVGYSAINSGYSQWGPGLKLFPAFWYIVHLYPELKGIMGR 1020
QY 1021 QNRTPTIVVWSVLASIFSLMVRIDPFSRTGTGDIIECGINC 1065
Db 1021 QNRTPTIVVWSVLASIFSLMVRIDPFSRTGTGDIIECGINC 1065

RESULT 2
US-09-900-237-30
Sequence 30, Application US/09900237
Patent No. US20020120124A1
GENERAL INFORMATION:
APPLICANT: Allen, Stephen
TITLE OF INVENTION: Plant Cellulose Synthases
FILE REFERENCE: BB1170 US CIP
CURRENT APPLICATION NUMBER: US/09/900, 237
CURRENT FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: 60/092, 844
PRIOR FILING DATE: 1998-07-14
PRIOR APPLICATION NUMBER: PCT/US99/15871
PRIOR FILING DATE: 1999-07-13

QY PRIOR APPLICATION NUMBER: 09/720383
QY PRIOR FILING DATE: 2000-12-21
QY NUMBER OF SEQ ID NOS: 33
QY SOFTWARE: Microsoft Office 97
QY SEQ ID NO 30
QY LENGTH: 1080
QY TYPE: PRT
QY ORGANISM: Trilicium aestivum
QY US-09-900-237-30
Query Match 79.4%; Score 4507; DB 10; Length 1080;
Best Local Similarity 78.4%; Pred. No. 0;
Matches 836; Conservative 98; Mismatches 107; Indels 26; Gaps 8;
QY 20 CQICSNVGTVDGDFRVAACDICSFVPCRCYERERKDNQSCPOCKTRYKRLKSPATP 79
Db 19 CQICADGLGTTLDGVDFTACDVCFRFYCRCYEHKEKEGTQALQCKTKRKRRGSPAIR 78
QY 80 GKDDEGLADEGTEVFENP-----OKERISRMGLMHLTRKGGEKMEGPQYDK----- 127
Db 79 GEEGDDTDDADDCGS-DENYPASCTEDOKOKIADRRMRSMNMTGSGGVGHKPYDSGEIGLS 137
QY 128 -----EVSNNHLRLRSRODTSGEFSASPERLSVSTIAGKRLPYSSDVNOSPRLIV 182
Db 138 KIDSGLIPROVSVYVINSQ-MSGELPGASPDHMHMSPGTGNISRRAPPY-VNHSPPSPRE 195
QY 183 DPGVGLGNVAMKERVDMKMKOEKNTGPV--STQAASE-NGVDIDASTDILADEALLND 238
Db 196 FSGSIGNVAMKERVDMKMKOEKNTGPV--STQAASE-NGVDIDASTDILADEALLND 255
QY 239 EAPQPSLRKVSIPSSSRINPYRMVIMRLVILCLFLHYRITNPVPAFALMLVSVICETIMF 298
Db 256 ETPQPSLRKVSIPSSSRINPYRMVIMRLVILCLFLHYRITNPVPAFALMLVSVICETIMF 315
QY 299 ALSMIDOPPKWFPVNRRETYLDRLALRYDREGESQOLAVIDIEVSTVDPLKEPPLVTANT 358
Db 316 ALSMIDOPPKWFPVNRRETYLDRLALRYDREGESQOLAVIDIEVSTVDPLKEPPLVTANT 375
QY 359 VLSILAVDYPVDKVCYFDDGAAMLSPESLAETSEFARKWPPCKKXSIIEPRAPEMVFAAK 418
Db 376 VLSILAVDYPVDKVCYFDDGAAMLSPESLAETSEFARKWPPCKKXSIIEPRAPEMVFAAK 435
QY 419 AKIDYLDKQVTSFVYKDRBRAMKREYEERKIRINALVSKALCPREGVWVMDGTPMPGNNT 478
Db 436 QKIDYLDKQVTSFVYKDRBRAMKREYEERKIRINALVSKALCPREGVWVMDGTPMPGNNT 495
QY 479 GDRPMIOVFILGQNGDLAEGNELPRLVYVSREKRPFGQHHKAGAMNALVRSVAVLTNGPF 538
Db 496 RDRPMIOVFILGQNGDLAEGNELPRLVYVSREKRPFGQHHKAGAMNALVRSVAVLTNGPF 555
QY 539 PFTLNDDCHYIINNSKALREAMCFIADBNILGKVQCYVOFPORFDGIDKNDRYANNTVFE 598
Db 556 QVTLNDDCHYIINNSKALREAMCFIADBNILGKVQCYVOFPORFDGIDKNDRYANNTVFE 615
QY 599 DYNLRLDGIQGVYVGTGVFNRTALYGBPPIKVKKHKKPSLSKLKSGSRKKSKAKKES 658
Db 616 DYNLRLDGIQGVYVGTGVFNRTALYGBPPIKVKKHKKPSLSKLKSGSRKKSKAKKES 673
QY 659 ESDKRSRHRDSTVPVNLDDIEEGVAGAGFDDKALLMSQSLKRRGOSAVVASTIME 718
Db 674 SSDKRSRHRDSTVPVNLDDIEEGVAGAGFDDKALLMSQSLKRRGOSAVVASTIME 733
QY 719 MENGVPSPATPENILKKAIVHISGVEDKSPMGMEIGMIVGVSVEEDILITGKKMARGRST 778
Db 734 MENGVPSPATPENILKKAIVHISGVEDKSPMGMEIGMIVGVSVEEDILITGKKMARGRST 793
QY 779 STYCPMLPAFKGAPINLSRLNQLVRLMAGSVEILFSRHCPIMVGYNGRLKFLERRAYVN 838
Db 794 STYCPMLPAFKGAPINLSRLNQLVRLMAGSVEILFSRHCPIMVGYNGRLKFLERRAYVN 853
QY 839 VNTTYPTTSTPLMAYCTLLAVCLFTNOFIIPQISINIASIWFLSLSTFATGILEMRSGV 898
Db 854 VNTTYPTTSTPLMAYCTLLAVCLFTNOFIIPQISINIASIWFLSLSTFATGILEMRSGV 913

Query	Match	Similarity	Score	DB	Length	Indels	Gaps
899	GVGIDEMWRNQNFGVAVIGVSAHLHFAVQGLIKVLVLAGIDNTNTYTSKASDEGDDAEYLTF	72.8%	413.5	10	881	5	3
914	GVGIDEMWRNQNFGVAVIGVSAHLHFAVQGLIKVLVLAGIDNTNTYTSKASDEGDDAEYLTF	72.8%	413.5	10	881	5	3
959	KWTLILPPTLLIVNLVGVAVGSAVYAINSGYSGMPLGALFAFWYIVLYPEFLGLM	72.8%	413.5	10	881	5	3
974	KWTLILPPTLLIVNLVGVAVGSAVYAINSGYSGMPLGALFAFWYIVLYPEFLGLM	72.8%	413.5	10	881	5	3
1019	GRQRRTPIVVMVSVLLASIFSLMVRIDPPTSRYVGPDIAGCINC	72.8%	413.5	10	881	5	3
1034	GRQRRTPIVVMVSVLLASIFSLMVRIDPPTSRYVGPDIAGCINC	72.8%	413.5	10	881	5	3
US-09-838-539-8	Sequence 8, Application US/09838539	72.8%	413.5	10	881	5	3
US-09-838-539-8	Patent No. US20020129401A1	72.8%	413.5	10	881	5	3
GENERAL INFORMATION:		72.8%	413.5	10	881	5	3
APPLICANT: Stalker, D. et al.		72.8%	413.5	10	881	5	3
TITLE OF INVENTION: Plant Cellulose Synthase and Promoter		72.8%	413.5	10	881	5	3
FILE REFERENCE: 15621/03/US		72.8%	413.5	10	881	5	3
CURRENT APPLICATION NUMBER: US/09/838,539		72.8%	413.5	10	881	5	3
PRIOR APPLICATION NUMBER: 2001-04-18		72.8%	413.5	10	881	5	3
PRIOR FILING DATE: 1996-10-29		72.8%	413.5	10	881	5	3
PRIOR APPLICATION NUMBER: 08/960,048		72.8%	413.5	10	881	5	3
PRIOR FILING DATE: 1997-10-29		72.8%	413.5	10	881	5	3
NUMBER OF SEQ ID NOS: 12		72.8%	413.5	10	881	5	3
SOFTWARE: FastSeq for Windows Version 4.0		72.8%	413.5	10	881	5	3
SEQ ID NO 8		72.8%	413.5	10	881	5	3
LENGTH: 881		72.8%	413.5	10	881	5	3
TYPE: PRT		72.8%	413.5	10	881	5	3
ORGANISM: Oryzae sativa		72.8%	413.5	10	881	5	3
US-09-838-539-8		72.8%	413.5	10	881	5	3
Query Match	72.8%	Score 413.5	DB 10	Length 881			
Best Local Similarity	86.8%	Pred. No. 0					
Matches 766	Conservative 55	Mismatches 56	Indels 5	Gaps 3			
188	GNVAMKEPVNDGMMKKQKQKENTGPRV---STQASSEKGV-DIDASTDILADEALLNDEARQP	72.8%	413.5	10	881	5	3
1	GNVAMKEPVNDGMMKKQKQKENTGPRV---STQASSEKGV-DIDASTDILADEALLNDEARQP	72.8%	413.5	10	881	5	3
244	LSRKYSIPSSRINPYRYMYIMLYLCLFHYRTINPYNAPNALMVLVSICEINFALSMI	72.8%	413.5	10	881	5	3
61	LSRKVPRLSSINPYRYMYIMLYLCLFHYRTINPYNAPNALMVLVSICEINFALSMI	72.8%	413.5	10	881	5	3
304	LDOFPKPEPVNRETYLDRLALRYDREGESPLAANDVIFSTVDPLKPEPLVTANTVLSIL	72.8%	413.5	10	881	5	3
121	IDQPKPMPINRETYLDRLALRYDREGESPLAANDVIFSTVDPLKPEPLVTANTVLSIL	72.8%	413.5	10	881	5	3
364	AVDPYVDKVCYVDDGAGMLSEFSLATSEFAKRWPFCKKYSIEEPAREMYTAAKIDY	72.8%	413.5	10	881	5	3
181	AVDPYVDKVCYVDDGAGMLSEFSLATSEFAKRWPFCKKYSIEEPAREMYTAAKIDY	72.8%	413.5	10	881	5	3
424	LKDVYQTSFVYDRAKMKREYEFIRIYALVSKALKCEDEGMVMDGFRWPGNNTGHDHG	72.8%	413.5	10	881	5	3
241	LKDVYQTSFVYDRAKMKREYEFIRIYALVSKALKCEDEGMVMDGFRWPGNNTGHDHG	72.8%	413.5	10	881	5	3
484	MIQVFLGNGSLDAEGNELPRLVYVYREKRPFGOHNNKAGAMNLLVRSAYVLNPGFLN	72.8%	413.5	10	881	5	3
301	MIQVFLGNGSLDAEGNELPRLVYVYREKRPFGOHNNKAGAMNLLVRSAYVLNPGFLN	72.8%	413.5	10	881	5	3
544	LDCHYIINNSALREAMCFMDPNLKGOCYVYQRFQDGIKNDRYANRNTVEFDINLR	72.8%	413.5	10	881	5	3
361	LDCHYIINNSALREAMCFMDPNLKGOCYVYQRFQDGIKNDRYANRNTVEFDINLR	72.8%	413.5	10	881	5	3
604	GLDGIQGVYVYGTGCVENRTALYGERPIYKHNKRPSSLSTLCGSSRKKNKAKKESQK	72.8%	413.5	10	881	5	3
421	GLDGIQGVYVYGTGCVENRTALYGERPIYKHNKRPSSLSTLCGSSRKKNKAKKESQK	72.8%	413.5	10	881	5	3

OY	664	KSGRHTSTVAVFNLNDTDEEVEAGPDEKALLMSQSLKREGQSAFVFASTLMENGG	722
Db	480	KSNNHVDVAVFNLNDEIEEGVAGGFDEKSLMSQSLKREGQSAFVASTLMETGG	539
OY	724	VPSATPENLLEKAIHVISCGEYEDKSDMGMEIGWYGSVTEEDILTGFMHARGMSIYCM	783
Db	540	VPSATPENLLEKAIHVISCGETEKTENGTEIGWYGSVTEEDILTGFMHARGMSIYCM	599
OY	784	PKLEPAFGSAPINLSDRLNOVLRNALGSVEILFSRHCPITWYGNRLKFLERFAYVNTTI	843
Db	600	PKRPAFGKSAPINLSDRLNOVLRNALGSVEILFSRHCPITWYGNRLKFLERFAYVNTTI	659
OY	844	YPLSISPLIAMCTLLAAGLFTNOFPIPOISINASIMWELSLSTISFATILERMRSVQID	903
Db	660	YPLSISPLIICVLPALICLLTGKFTLPIEISNPASTWFLSTISFATILERMRSVQID	719
OY	904	EMMRNEQFWVIGVSAHLFAVFOGILKVLAGIDNTFTSKASDEDDGFAELYLFKMTTL	963
Db	720	EMMRNEQFWVIGVSAHLFAVFOGILKVLAGIDNTFTSKASDEDDGFAELYLFKMTTL	779
OY	964	LIPPTLLILNLCGVYAVGVASTAINSGVSWGPLGKLFEPATWVYVHLPLFLKGLMRONR	1023
Db	780	LIPPTLLIILNLCGVYAVGVASTAINSGVSWGPLGKLFEPATWVYVHLPLFLKGLMRONR	839
OY	1024	TPTIYVWAVLLASIFSLLMWIRIDPFTSRVGPDTLEGGINC	1065
Db	840	TPTIYVWAVLLASIFSLLMWIRIDPFTSRVGPDTGTGGINC	881
RESULT 4			
US-09-900-237-26			
; Sequence 26, Application US/09900237			
; Patent No. US20020120124A1			
; GENERAL INFORMATION:			
; APPLICANT: Allen, Stephen			
; TITLE OF INVENTION: Plant Cellulose Synthases			
; FILE REFERENCE: B81170 US CIP			
; CURRENT APPLICATION NUMBER: US/09/900,237			
; CURRENT FILING DATE: 2001-07-06			
; PRIOR APPLICATION NUMBER: 60/092,844			
; PRIOR FILING DATE: 1998-07-14			
; PRIOR APPLICATION NUMBER: PCT/US99/15871			
; PRIOR FILING DATE: 1999-07-13			
; PRIOR APPLICATION NUMBER: 09/720383			
; PRIOR FILING DATE: 2000-12-21			
; NUMBER OF SEQ ID NOS: 33			
; SOFTWARE: Microsoft Office 97			
; SEQ ID NO 26			
; LENGTH: 1091			
; TYPE: PRT/			
; ORGANISM: Impatiens balsamita			
US-09-900-237-26			
Query Match 70.5%; Score 3999.5; DB 10; Length 1091;			
Best Local Similarity 69.5%; Pred. NO. 0;			
Matches 750; Conservative 126; Mismatches 162; Indels 41; Gaps 14;			
OY	10	KPKMNIYQTOQICSDNAGTKVVDGRFACIDCSPVPCPYEYERAKGNSCPQCKTRY	69
Db	31	KPLKAMNOQIOICEDYVKGSAITGDTFVACNCCGPPVPCPYEYERKKGNOCCPQCKTRY	90
OY	70	KRLKSPAIIPGKDDGLADGTVGFNFYPOREKISERMLGWLHTGKGEEMGEPOYDKREV	129
Db	91	KRKQKSPYVEDDEEDV-DLLENFNFSGKCKNKK-----VTTARRPWQD-QODIEL	143
OY	130	S-----HNH-----LPRLTSRDSTGSEFSAASPERLVSSTIAG-----GKRLPYSSDVNS	176
Db	144	SVSSSRHDSQOQPVLLTHGHSVSGEI--PTPDNHSIRTSQIPGPEKSIPIYIDRPQV	201
OY	177	PNRRIVDP-----VGLGNVAMKEKVDYDKMKKQENKTPVSTQASSEGGYDIDASTIOL	230
Db	202	AVRIIVDSKDLNSYGLGAVDMKEKREVEGKWLQEKNNVMQMSRYPEGG--DTEGTGSG	259

```

01 231 ADELLINDEAOPSRKSISSSRNPRAWIMLRATVLCFJHFRINPNPMAFALY 290
02 260 EELOMADDDIQMSRIYPISTHILTPRVVYLLILLLIGFLQTRCHTPVKADAPLWLT 319
03 291 SVICEIWEFALSMIIIDQEPKMEPVNRETYLDRILATRYDREGEPSOLAADIFVSTVDPLKE 350
04 320 SVICEVWFALSMIIIDQEPKMSPVNRETYLDRILSMNFORGEPSOLAPIDFVSTVDPLKE 379
05 351 PPLVTANTVLSILAVDYVDKVCYVFDDGAAMISFESLAETSEFAKWPDPCKKYSTIEP 410
06 380 PPLVTANTVLSILAVDYVDKVCYVSDGGSAMLTFFELSETAEFAKWPDPCKKHSIEDP 439
07 411 RAPEKYEFAKIDYLDKDKVQTSFVDRBRAMKREYESEKIRINAYLSKALCKPEEGVMYMDG 470
08 440 RAPEYEFQKIDYLDKDKVQTSFVDRBRAMKREYESEKIRINAYLSKALCKPEEGVMYMDG 499
09 471 TPWPGNNITGDHPGMIOVFLGONGGLDAGNELPRLYVYSREKRPFGQHHKAGANALVR 530
10 500 TPWPGNNISRDHPGMIOVFLGHSGGDDTSGNELPRLYVYSREKRPFGQHHKAGANALIR 559
11 531 VSAVLTPNGPILNDCDHY INNSKALREAPCFLMDPNIGKOVYQVQFQOREDGIDKNDRY 590
12 560 VSAVLTPNGVLLNDCDHYFNNSCKLKAEMCFMDPNIGKKTQCYQVQFQOREDGIDLHRY 619
13 591 ANRNTVPEFDILRGIDIGQPVYVTCGVFNRTALYGEPRPLKYNHKKRPSLLSKCGSR 650
14 620 ANRNTVPEFDILNKGIDIGQPVYVTCGCCFNQALYGDVPLTEDELPNIIKSCCSR 679
15 651 KNSKA-KESDEDKKSGRHTDSTJVPFNLDIEEGVEGAGFDEKALMSOMSKLRGQ 709
16 680 KKGKGGNKKYIDKNRALKRTESTAPIFNMEDIEBIE--GYDDESFYMAO-SYKRRGQ 736
17 710 SAVEFASITLMEGVGPSATPENILKEAIIHVISCGYEDKSDWMEIGMITYGSVEDILTG 769
18 737 SPVLIAATFMOGQGLPSTNSATILKEAIIHVISCGYEDKTEWKEIGMITYGSVEDILTG 796
19 770 FKMAHARGRSITCYMPKLPFAFGSAPINSDRLNOYLKRALSGVELLESRHPICYWYGR 829
20 797 FKMHRGMSITCYMPRPAPFGSAPINSDRLNOYLKRALSGSIELILSRHCPITWYGSGR 856
21 830 LKLEFERFVYVNTTYPIPISTPLMWCTLAVCLFNOPLIQOISNIAISWFLSTLSFA 889
22 857 LKLEFERLAVNTIYIPLISIPLAYCTIPALICLLTGKTIPEISNIAISWFLSTLSFS 916
23 890 TGLEMRMSGVGIDEMWNRNEQFWYIGVSAHLFAVQGLILKYLAGIDINFTYTSKASDED 949
24 917 TGLELRMSGVTLDEDMWRNEQFWYIGVSAHLFAVQGLILKYLAGIDINFTYTSKASDED 976
25 950 GDFAEFLYFKNTTLLIPTLLILYNLVGAVGSAINSIGTOSWGPLGKLEFAFWIVH 1005
26 977 GDFAEFLYFKNTSLIDPPTLLILYNLVGAVGSAINSIGTOSWGPLGKLEFAFWIVH 1036
27 1010 LYPFLKGLMGONRPTIIVVWVSVLASIFSLWIRIDPFS---RVNGPDLEGGICN 1065
28 1037 LYPFLKGLMGONRPTIIVVWVSVLASIFSLWIRIDPFSYDSTKRG---QCGIDC 1091
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
7
```

```

; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 1148
; TYPE: PRT
; ORGANISM: Zea mays
US-09-900-237-4

Query Match      69.1%; Score 3922; DB 10; Length 1148;
Best Local Similarity 67.2%; Pred. No. 0;
Matches 734; Conservative 141; Mismatches 161; Indels 56; Gaps 17

QY      4 EGETAGKPMKNTVPQTCOICSDNNGKTVYGDGFVACDICSFPCRCRCYERERDGNOSCP 63
Db      82 DGEPPKPPDQORNGVCCQICGDDVGNPNPGEFPVACAFPCRCRCYEREREGTQNC 141
QY      64 QCKTRKRLKSGPALPGDKDEGLADEGTVEFENYPOK---EKISERMGLMHLTRGGEEM 120
Db      142 QCKTRFKRLKGCARYPGDEEDGV--DLENEFNMSDKHSQYLAESEMLAHMSYGGADL 200
QY      121 -GEOPYDKEVSHNHLPRILTSRO---DTSGEFSASPERLSVSTTAGKR--LPYSSVD 173
Db      201 DGVPPQFHPHP--NVPLLTNGQMVNDIPDPQHIALVP-----SFVGGGGRINHPLP--ADP 252
QY      174 NQSPKRRITVP-----VGLGNVAKKERDQKMKMKQCKNTGVSPTQASBERGVDDIAS 227
Db      253 NLPVQPRSDPDKDLAAYGVGSVANKERMSKQOKER-----MNOTRNDGDDGDDDD-- 305
QY      228 DLADEALLNDARQPLSRKVSIPSSRIPIRYMVLRLVLIICLFHYRTITPNVPAPAL 287
Db      306 ---ADLPPLM--DKARQPLSKITLPLSQINPYRMIIITIRLVLCFFHYHYVMHNPAPAL 361
QY      288 MLVSYVICETWFLSMILDOFPKMFPPVNRETYLDRLALRYDRREGSPSOLAVIDIYSTVD 347
Db      362 MLISYICETWFLSMILDOFPKMFPIERTETYLDRLSLRDKEGHPSQLAPVDFEYSTVD 421
QY      348 LKEPPLVANTVYLSLAVDYPVDKYSCTYFDDGAMLSESLAETSEFARKVYRFECKYS 407
Db      422 LKEPPLVANTVYLSLAVDYPVDKYSCTYFDDGAMLSESLAETSEFARKVYRFECKYS 481
QY      408 IERRAPENWEFAKIDYLDKQVOTSFVKDRAKREVEERKIRINALVYSALKCREGVMV 467
Db      482 LERRAPENWFOOKIDYLDKQVAPNVEYREKRAKREVEEKVIRINALVYAKQKREGEWY 541
QY      468 QDGTWMPGNNTGDHGMQIOVFLGONGGLDAEGNELPRLYVVSNEKRPGEONHKKAGAMA 527
Db      542 QDGTWMPGNNVNDHGMQIOVFLGOSGHDVEGELPRLYVVSNEKRPGEONHKKAGAMA 601
QY      528 LVRSASVLTNGEFLINDCDCHYINNASKALREAMCFIAMDNLGQVSYVOREPQFSDIKN 587
Db      602 LVRSASVLTNAYVLTINDCDCHYINNASKALKEAMCFEMDNLGQVSYVOREPQFSDIDRH 661
QY      588 DRYARNVVEPQINLRGIDLGIGPYVVGSCFNFRLALGYEERIKVKKKPS----- 640
Db      662 DRYARNVVEPQINNRGIDLGIGPITVIGCYVFRQALGLYDLP---KTKKPPSRPCNC 718
QY      641 ---LLSKLQKSGARKKNSKAKKESDK---KSGRHNDSTVAPFNLLDIEGVEGAFDDE 693
Db      719 PKWICCCCFGRKRRKKTKTKTSKPRFEKIKKLFEKKKENQAPAYALGEIDEAPGA--ENN 776
QY      694 KALMSQMSLEKRFQSOAVFAVSTLMENGSGVPPSATPENULKEALHIVISCGYEDKSDKMG 753
Db      777 KASIVNQOKLEKKFQOSSVFAVSTLLENGSTLKSASPALEALHIVISCGYEDKSGWGC 836
QY      754 ELGATVYGSVTEDILTGFKMHAAGMSTIYCMPLPAPFKGSAPINLSRLNOVLRYMALGSVE 813
Db      837 DIGWITVGSVTEDILTGFKMHGGMSTIYCIPIPRAAFKGSAPINLSRLNOVLRYMALGSJE 896
QY      814 ILFSHRCPIWGYNGRKLFLERFAAVNTIYVITSPILMYCCTILAVCFETNPQIIPQIS 873
Db      897 IFSNHRCPIWGYGGGKLFLERFSTINSIVYWTISPIILAYCTIPACILDLTGKFTPELN 956
QY      874 NIASIMPLSLFLSIPATGILLEMRSVGVDIDEMWRNEDQVNVIGVSAHLEPAVFOGILKVL 933

```

Db 957 NVASLWMSLFCIFAFSILEMRWSGVGIDMMREOFVWIGVSSHLFAVFOGLKLVIA 1016
Qy 934 GIDNFWTSKASDEDDGFAELYLFWKTTLLIPPTLLIIVNLGVVAIVSAINSQYSW 993
Db 1017 GVDTSFVTSKGGD-DEESELTYFKWTTLLIPPTLLIIVNLGVVAIVSAINSQYSW 1075
Qy 994 GFLGKLEFARVIVHLXPLFKGLMGRONRPTIIVVWSVLASIFSLMWRIDPFTSRV 1053
Db 1076 GFLGKLEFARVIVHLXPLFKGLMGRONRPTIIVVWSVLASIFSLMWRIDPFLAKD 1135
Qy 1054 TGPDILECGINC 1065
Db 1136 DGPLLEECGLDC 1147

RESULT 6
US-09-900-237-10
; Sequence 10, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen
; TITLE OF INVENTION: Plant Cellulose Synthases
; FILE REFERENCE: Bb1170 US CIP
; CURRENT APPLICATION NUMBER: US/09/900, 237
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: 60/092,844
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: PCT/US99/15871
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: 09/720383
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
; LENGTH: 1086
; TYPE: PRT
; ORGANISM: Zea mays
US-09-900-237-10

Query Match 66.3%; Score 3875.5; DB 10; Length 1086;
Best Local Similarity 66.7%; Pred. No. 0;
Matches 728; Conservative 145; Mismatches 161; Indels 57; Gaps 18;

Qy 4 EGETAGPKMKIVIVQTOICSDNIGKTVGDRFAACDICSFPVRCPEYERKGNOSCP 63
Db 23 DGDGPPRPQRQNGOVCOICGDDVGLAPGDPFVACNCAFPVCDCEYERREGTNC 82
Qy 64 QCKTRYKRLKSPAIIPGDKDEDGLADECTVEFNYP--QEKISERMLGMHLTRGK-GEEM 120
Db 83 QCKTRYKRLKCCQVNTGDEEDGY-DLDNDEFNMDGHSQVASEMLIGHMSYGRGDPN 141
Qy 121 GEPQYDEVSHNLPLRLTSRQ--DTSGEFSASPERLSVSTIAGKR--LPPSSDN 174
Db 142 GAPAQFO--LNPVPLLTNGQWVDIPPEQHALVPSFG-----GGKRHPHLPY-ADPS 193
Qy 175 QSPRRRIYDP-----VGLGNVANKERVDGKMKQEKTKGVSVOAASRECGVIDASTD 228
Db 194 LPVQPRMDPSKDLAAYGVGSVANKEREMNKQKER-----MQGTGNDGGDDDD-- 245
Qy 229 ILADEALINDAROPILSRKVSIPSSRIINPYRMVJMLRVITCLFLHYRTNPPVPAFLM 288
Db 246 --AALPLM-DEARQOLSKRIPLPSSQINPRMIIIRLVGLGFFHYVMHPVNDAPFLM 302
Qy 289 LVSVICELWALSLILDOFPKWPFPVNRRTYLDRLALRYDRGEPQSOLAVIDFVSTVPL 348
Db 303 LISVICEIWFAMSWILDOFPKWPPIERETYLRLSLRDKGQSPQLAPIDFVSTVPL 362
Qy 349 KEPPLVANTVLSILAVDYPVDKSCYVFDGAMLSRESLAHSEFRAKVPFCKKYSI 408
Db 363 KEPPLVNTVLSILSDVYPVDKSCYVSDGAMLSFEALSETSEFAKVPFCKKYSI 422
Qy 409 EPRAPENYFAAKIDYLDKGVQTSFVKDRRAKREYEEFKIRINALYSKALCPEEGVWQ 468

Db 423 EPRAPENYFOOKIDYLDKDVANAFVERRRAKREYEEFKIRINALYAKAKVPEEGVWQ 482
Qy 469 DGTIPKGNNTGDHRCMIQVFLQNGGLAEGNELPLRLVYVSEKRPQRQHHKKAAMNL 528
Db 483 DGTIPKGNVNRDHPQMIQVFLQSGGLDCGEGNELPLRLVYVSEKRPQYHHKKAAMNL 542
Qy 529 VRVSAVLNPGFPIILDCDHYIINNSKALREAMCFMADNGLKOVGYVOPQRFSDIDKND 588
Db 543 VRVSAVLNAPILNLDGDHIIINNSKALREAMCFMADNGLKOVGYVOPQRFSDIDKND 602
Qy 589 RYANRNVFEDINLGLDGIQGVYVTCGVFNRTALYGEPPIRVKKHKSLSLKG- 647
Db 603 RYANRNVFEDINLGLDGIQGVYVTCGVFNRTALYGEPPIRVKKHKSLSLKG- 657
Qy 648 -----GSRKKNKAKKESDKR--SGRHTDSVYPVFNLDIDIEGVEGAFDDEK 694
Db 658 WPKWCFCCCGENRRQKTKRPTKTKKLLFEKKEENOSPVALGEIDEAIPA--ENEK 715
Qy 695 ALLMSOMSLERFGOSAVFVASTLMENGVPSPATPENLKEAIIHVISCYEDKSDMGME 754
Db 716 AGIVNQQLEKKFGQSSVFTVSTLENGGTLKSASPASLKEAIIHVISCYEDKSDMGME 775
Qy 755 IGMIVSVTEDILTGFKMHARGWSIYCPKLPARFGSAPINLSDRLNOVLRMALGSVEI 814
Db 776 IGMIVSVTEDILTGFKMHARGWSIYCPKLPARFGSAPINLSDRLNOVLRMALGSVEI 835
Qy 815 LFSRCPITWYNGNLKLERFAYVNTTITPTISPLMYCTLAVLCFTNOFTIPIQISN 874
Db 836 FFSNHCPLWYGGGLKFLERSYINSIVPTTSIPLAYCTLPACILCTGKFTIPELNN 895
Qy 875 IASIVFLSLPISFATGILFEMRWSGVIGIDMMRNQFVWIGVSHLFAVFOGLKLVIA 934
Db 896 VASLWMSLFCIFAFSILEMRWSGVGIDMMRNQFVWIGVSHLFAVFOGLKLVIA 955
Qy 935 IDNFWTSKASDEDDGFAELYLFWKTTLLIPPTLLIIVNLGVVAIVSAINSQYSW 994
Db 956 VDSFTVTSKGGD-DEESELTYFKWTTLLIPPTLLIIVNLGVVAIVSAINSQYSW 1014
Qy 995 PFLGKLEFARVIVHLXPLFKGLMGRONRPTIIVVWSVLASIFSLMWRIDPFTSRV 1054
Db 1015 PFLGKLEFARVIVHLXPLFKGLMGRONRPTIIVVWSVLASIFSLMWRIDPFLAKD 1074
Qy 1055 GPDILECGINC 1065
Db 1075 DGPLLEECGLDC 1085

RESULT 7
US-09-900-237-8
; Sequence 8, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen
; TITLE OF INVENTION: Plant Cellulose Synthases
; FILE REFERENCE: Bb1170 US CIP
; CURRENT APPLICATION NUMBER: US/09/900, 237
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: 60/092,844
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: PCT/US99/15871
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: 09/720383
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 1165
; TYPE: PRT
; ORGANISM: Zea mays
US-09-900-237-8

Query Match 67.7%; Score 3843.5; DB 10; Length 1165;

[illegible]

	Query Match	Similarity	62.1%;	Score	3473.5;	DB	10;	Length	974;
Best Local	Similarity	62.1%;	Score	3473.5;	DB	10;	Length	974;	
Matches	660;	Conservative	129;	Mismatches	176;	Indels	97;	Gaps	15

72 LKSPALIPDGDDEGLADGCTVEFNPPOKEKISERNLGHHLTRGKEGEBEPDYDKEVSH 131
55 -GSP-----YDENLDD-----VERATGDQSTMAHL--NKSDVG-----HIA 90
132 NMLPRLTSHODTSGEESAASPERLSVSTIAGGRKRLPYSDVQSPNRRIYDPVGLGNA 191
91 RHISVSSTJLDSMAEEN-----GNSI 111
192 MKERYDGMKMOEKNTGPVSTOASERGVDDIDASTDILLADEALLND---EAROPLSRK 247
112 MKRWESMEKKNKKKKPATTKYERAE-----IPPEQOMEDKAPDASQCLSTI 161
248 VSPSSRINPRMVMIMLRVILICLPHVITNPVPMAPALMVLVSCIEIPALSTLDDP 307
162 IPIPSRLAPRYTVIIMRLILGLFPHYRNPVDSAFGLMTVSCIEIPAFSWLDDP 221
308 PKMFVNPRTYDLRLARTDREGESQLAADVIFVSTVPLKEPPLVANTVLSIADY 367
222 PKMYPNRTTYIDRLSAREGEDEPLAADVDFVSTVPLKEPPLITANTVLSIADY 281
368 PVDKVSCTVDDGAAMLSESLAETSEFARKWVPCKKYSIEPRAPEMYFAAKIDYLDK 427
282 PVDKVSCTVDDGAAMLSESLAETSEFARKWVPCKKYSIEPRAPEMYFAAKIDYLDK 341
428 VQTSFYKDBRMRKREYEEKIRINLALYSKALCPEEGWMOGSTPMRGANTGDHPGMIOY 487
342 VQTSFYKDBRMRKREYEEKIRINLALYSKALCPEEGWMOGSTPMRGANTGDHPGMIOY 401
488 FFGONGGLDAEGNELPRLVYVSEKRPGRFQHHKAGAMALVSVAVLTNGFPIILDD 547
402 FFGONGGLDAEGNELPRLVYVSEKRPGRFQHHKAGAMALVSVAVLTNGFPIILDD 461
548 HTNNSKALREAMCEFLMDPILKQVCYVOPORFDGIDKNDRYANNTVFEDINLGLDG 607
462 HTNNSKALREAMCEFLMDPILKQVCYVOPORFDGIDKNDRYANNTVFEDINLGLDG 521
608 IGGPVYVGTGVCFNRTALYGEPRITKVRKKPSLILSKLGGSKKKSKAKKSDKKSR 667
522 IGGPVYVGTGVCFNRTALYGEPRITKVRKKPSLILSKLGGSKKKSKAKKSDKKSR 579
668 HTDSTVPVNLDDIEBEGVAGFDD-EKALIMSOMSLERFGOSAVFVASTLMENGVP 726
580 REIDLAIFNLREID-----NYDEYERMLISQTSFEKTFGLSSVFISTLMENGVAE 633
727 SATPENLKEALHIVISCGYEDKSDWMEIGWITGVSTEDILTGFKMHARGMSIYCMPL 786
634 SATPENLKEALHIVISCGYEDKSDWMEIGWITGVSTEDILTGFKMHARGMSIYCMPL 693
787 PAFKSGAPINLSDRLNOVLRLMALGSVEILFSRHCPITWGY-NGRLKFLERFAYVNTIYP 845
694 PAFKSGAPINLSDRLNOVLRLMALGSVEILFSRHCPITWGY-NGRLKFLERFAYVNTIYP 753
846 IASPIILMYCTLLAVCLFTNOPIIPIQISNIASIMFLSLFISFATGILMRMSGVIGDM 905
754 FLSPLIANCSLPAICLLTGKFLIPTLSLAVSLFGLFSLITVAVLELRMSGVIGDM 813
906 MNEQFVIGVSAHLFAVFOGILKVLADIDNFTVTSKASDEGDFAELYLEFKWTTLLI 965
814 MNEQFVIGVSAHLFAVFOGILKVLADIDNFTVTSKASDEGDFAELYLEFKWTTLLI 872
966 PPTTLLIVNLGVVGVSAINSQOSMGPLFGKLFARFVAVIHLVLPFLKGLMRGNRP 1025
873 PPTTLLIVNLGVVGVSAINSQOSMGPLFGKLFARFVAVIHLVLPFLKGLMRGNRP 932
1026 TIVVWVSVLASIFSLMWRIDPFTSRVYGPDLLE--CGINC 1065
933 TIVVWVSVLASIFSLMWRIDPFTSRVYGPDLLE--CGINC 974

RESULT 11
US-09-900-237-18
; Sequence 18, Application US/09900237
; Patent No. US2002012012A1

GENERAL INFORMATION:
; APPLICANT: Allen, Stephen
; TITLE OF INVENTION: Plant Cellulose Synthases
; FILE REFERENCE: B01170 US CIP
; CURRENT APPLICATION NUMBER: US/09/900,237
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: 60/092,844
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: PCT/US99/15871
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: 09/720383
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 18
; LENGTH: 793
; TYPE: PRT
; ORGANISM: Glycine max
US-09-900-237-18

Query Match 56.6%; Score 3216; DB 10; Length 793;
Best Local Similarity 72.3%; Pred. No. 4,8e-268;
Matches 573; Conservative 118; Mismatches 90; Indels 12; Gaps 5;

279 NPYNAPALMVLVSCIEIPALSTLDDPCKPVPVRETYDLRLARTDREGESQLA 338
4 HPYNDAYGLMLTSVIEIEMFAVSWIDDPKWPPIORETYDLRLSRYKEKPSLSV 63
339 DIFVSTVPLKEPPLITANTVLSIADYVPOVKSVCYVDDGAAMLSESLAETSEFARK 398
64 DIFVSTVPLKEPPLITANTVLSIADYVPOVKSVCYVDDGAAMLSESLAETSEFARK 123
399 WPFCKKYSIEPRAPEMYFAAKIDYLDKQVTSFYKDBRMRKREYEEKIRINLALYSKAL 458
124 WPFCKKYSIEPRAPEMYFAAKIDYLDKQVTSFYKDBRMRKREYEEKIRINLALYSKAL 183
459 KCPBEGWMOGSTPMRGANTGDHPGMIOYFLOONGGLDAEGNELPRLVYVSEKRPGRF 518
184 KCPBEGWMOGSTPMRGANTGDHPGMIOYFLOONGGLDAEGNELPRLVYVSEKRPGRF 243
519 HKKAGAMALVSVAVLTNGFPIILDDCHYINNSKALREAMCEFLMDPILKQVCYVOP 578
244 HKKAGAMALVSVAVLTNGFPIILDDCHYINNSKALREAMCEFLMDPILKQVCYVOP 303
579 QRFDGIDKNDRYANNTVFEDINLGLDGIOGPVYVGTGVCFNRTALYGEPRITKVRKK 638
304 QRFDGIDKNDRYANNTVFEDINLGLDGIOGPVYVGTGVCFNRTALYGEPRITKVRKK 363
639 -----PILLSKLGSSKSKAKKESKSGRTSDTVFVNLDIEBEGVAGFDD 692
364 KTCNCMPKWCCLCCGSRKKRNANSKE--KKRKVHSEASKOIHALENTAENEGT--NN 419
693 EKALIMSOMSLERFGOSAVFVASTLMENGVPASATPENLKEALHIVISCGYEDKSDW 752
420 EKTSNLTQKLEKRGOSVYFVASTLDDGVPFGVSPASLKEALQVISCYEDKTEWG 479
753 MEIGWITGVSTEDILTGFKMHARGMSIYCMPLPAFKSGAPINLSDRLNOVLRLMALGSV 812
480 KEVGMWITGVSTEDILTGFKMHARGMSIYCMPLPAFKSGAPINLSDRLNOVLRLMALGSV 539
813 ELFSRHCPITWGYNGRLKFLERFAYVNTIPIISIPILMYCTLLAVCLFTNOPIIPI 872
540 ELFSRHCPITWGYNGRLKFLERFAYVNTIPIISIPILMYCTLLAVCLFTNOPIIPI 599
873 SNIASIMFLSLFISFATGILMRMSGVIGDMWNEQFVIGVSAHLFAVFOGILKVL 932
600 SNIASIMFLSLFISFATGILMRMSGVIGDMWNEQFVIGVSAHLFAVFOGILKVL 659
933 AGIDNFTVTSKASDEGDFAELYLEFKWTTLLIIPPTTLLIVNLGVVGVSAINSQOS 992
660 AGVNTNFVTSKASD--DGEFSELYIFKWTSLIIPPTTLLIINNIVGVVAGISDAINN 718
993 WGPLFGKLFARFVAVIHLVLPFLKGLMRGNRPITIVVWVSVLASIFSLMWRIDPFTSR 1052

Db 719 WGPLEGRLEFALMVLHLYPFLKGLGKQDRMPITLLVMSILLSLTLFWVRINPVSFR 778
QY 1053 VTGPDLIEGICINC 1065
Db 779 -DGPVLEICGLNC 790

RESULT 12
US-09-900-237-24
; Sequence 24, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen
; TITLE OF INVENTION: Plant Cellulose Synthases
; FILE REFERENCE: Bb1170 US CIP
; CURRENT APPLICATION NUMBER: US/09/900,237
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: 60/092,844
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: PCT/US99/15871
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: 09/720383
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Plectramnia pentandra
US-09-900-237-24

Query Match
Best Local Similarity 52.9%; Score 3003.5; DB 10; Length 740;
Matches 547; Conservative 94; Mismatches 90; Indels 13; Gaps 5;

QY 338 RGEPSOLAANDIFVSTVDPLKEPPLVTANTVLSILAVDYPVDCVSCYVFDGAAMLSPFE 387
Db 3 KEGKPSLAGIDIFVSTVDPMKEPPLITANTVLSILAVDYPVDCVSCYVFDGAAMLSPFE 62

QY 388 SLAESEFARKKVPCKKXISTEPRAPEWYFAKIDYLDKDVOTSFVKDRAMKREYEERK 447
Db 63 ALSESEFARKKVPCKKXISTEPRAPEWYFAKIDYLDKDVOTSFVKDRAMKREYEERK 122

QY 448 IRINLYSALKCEPREGVMDGTPMPGNNTGDPHGMIOVFLGONGGDAEGNELPRLVY 507
Db 123 VRINLYSALKCEPREGVMDGTPMPGNNTGDPHGMIOVFLGONGGDAEGNELPRLVY 182

QY 508 VSREKRRPGFOHKKAGANALVRSAVLTNGPFIINLDCDHYINNSKALREAMCFMLDPEN 567
Db 183 VSREKRRPGFOHKKAGANALVRSAVLTNGPFIINLDCDHYINNSKALREAMCFMLDPEN 242

QY 568 LGKQVCYVOPQOREGIDKNDRYANRNVFFDINLRGIDGIDGPPYVGTGCVFNRTALYG 627
Db 243 SGKKLVCYVOPQOREGIDKNDRYANRNVFFDINLRGIDGIDGPPYVGTGCVFNRTALYG 302

QY 628 YEPPIKVVHKK-----PSLSKLCGSGRKKNSKAKKESDKKSRHDSITVPVNLDOI 681
Db 303 YDAPVTKKSPGACACWPKMLCCCGGSRKKNSKAKKESDKKSRHDSITVPVNLDOI 359

QY 682 BEGVGAGFDEKALLMSOMSLERFGOSAVFVASTLMENGVPSPAPENILKEAIIHY 741
Db 360 BEGM--GGINSEKSETTLPKLEKKFGSPVFASTLLEDGVPDAPATALLKEIYI 417

QY 742 SCGIDKSDMGMEIGMIVGSYVEDILITGFKMHARCRSTYCMPLKPAFGSAPINLSRDL 801
Db 418 SCGIDKSDMGMEIGMIVGSYVEDILITGFKMHARCRSTYCMPLKPAFGSAPINLSRDL 477

QY 802 NOVLEWALGSVILSRHCPIWYGNRGLKLEFAVYNTTITPITSIPLLWYCTLLAY 861
Db 478 NOVLEWALGSVILSRHCPIWYGNRGLKLEFAVYNTTITPITSIPLLWYCTLLAY 537

QY 862 LETNOFIIPOISINIASIMFSLSLISIFATGILEMRSGVIGIDEMWNRNEOFNIVGVSANL 921

Db 538 LITGFIYVEISINVASIILEMFLFIATNSILEMVGQGVGIDMWRNEOFNIVGVSANL 597
QY 922 FAVFOGILKVLGIDITNFTVTKASDEDDDEAFELYFKKTTLLIPTTLLIYNLVGVVAG 981
Db 598 FALFOGLKVLGAVNTNFTVTKADE--GDFSELVFKKTTLLIPTTLLIYNLVGVVAG 656

QY 982 VSAINSYGOSWCPLEKLEFAFWYVHLHYPLKGLMGQRNRPRTIVVWVSYLLASISFL 1041
Db 657 VSDALINGYDSNGPLKLEFAFWYVHLHYPLKGLMGQRNRPRTIVVWVSYLLASISFL 716

QY 1042 LWRIDPFTSRVTPDLIEGICINC 1065
Db 717 LWRINPVSFR--DGPVLEICGLNC 739

RESULT 13
US-09-900-237-31
; Sequence 31, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen
; TITLE OF INVENTION: Plant Cellulose Synthases
; FILE REFERENCE: Bb1170 US CIP
; CURRENT APPLICATION NUMBER: US/09/900,237
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: 60/092,844
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: PCT/US99/15871
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: 09/720383
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 31
; LENGTH: 685
; TYPE: PRT
; ORGANISM: Gossypium hirsutum
US-09-900-237-31

Query Match
Best Local Similarity 49.5%; Score 2808; DB 10; Length 685;
Matches 512; Conservative 77; Mismatches 77; Indels 22; Gaps 6;

QY 397 RKMVPCKKXISTEPRAPEWYFAKIDYLDKDVOTSFVKDRAMKREYEERKIRINLVSK 456
Db 1 RKMVPCKKXISTEPRAPEWYFAKIDYLDKDVOTSFVKDRAMKREYEERKIRINLVSK 60

QY 457 ALKCEPREGVMDGTPMPGNNTGDPHGMIOVFLGONGGDAEGNELPRLVYVSREKRRPG 516
Db 61 ALKCEPREGVMDGTPMPGNNTGDPHGMIOVFLGONGGDAEGNELPRLVYVSREKRRPG 120

QY 517 OHKKAGANALVRSAVLTNGPFIINLDCDHYINNSKALREAMCFMLDPNLKQVCYVQ 576
Db 121 OHKKAGANALVRSAVLTNGPFIINLDCDHYINNSKALREAMCFMLDPNLKQVCYVQ 180

QY 577 FQOREGIDKNDRYANRNVFFDINLRGIDGIDGPPYVGTGCVFNRTALYGEPPIKVVH 636
Db 181 FQOREGIDKNDRYANRNVFFDINLRGIDGIDGPPYVGTGCVFNRTALYGEPPIKVVH 240

QY 637 KR-----PSLSKLCGSGRKKNSKAKKESDKKSGR--TSTVVPVNL 678
Db 241 KR-----PSLSKLCGSGRKKNSKAKKESDKKSGR--TSTVVPVNL 300

QY 679 DDEGVGAGFDD--EKALLMSOMSLERFGOSAVFVASTLMENGVPSPAPENILKEA 737
Db 301 DEGVGAGFDD--EKALLMSOMSLERFGOSAVFVASTLMENGVPSPAPENILKEA 358

QY 738 HVISGCVSDKSDMGMEIGMIVGSYVEDILITGFKMHARCRSTYCMPLKPAFGSAPINL 797
Db 359 HVISGCVSDKSDMGMEIGMIVGSYVEDILITGFKMHARCRSTYCMPLKPAFGSAPINL 418

QY 798 SDRLNOVLWALGSVILSRHCPIWYGNRGLKLEFAVYNTTITPITSIPLLWYCTLL 857

Db 419 SRRHCVLRKMGALGSVEIFLSRHCPIMYGGKGLKMLERLAYINTIYPPFSIPLACYI 478
QY 858 LANCETNOFIIPQISNIASIWFLSLFISFATGILEMRSGVIGIDEMMRNEQFVIGY 917
Db 479 PAVCLITGFIIPITLSNLTSVWFLALFSLIATGVLRLMSGSVIGIDEMMRNEQFVIGY 538
QY 918 SAHLFAVFOGILKVLAGIDNTFTVTSKASDEGDAELYLKFWTLLIPPTLLIPLVNLV 977
Db 539 SAHLFAVFOGILKVLAGIDNTFTVTSKASDEGDAELYLKFWTLLIPPTLLIPLVNLV 597
QY 978 VVAGSVYAINSGYSGMPLFGKLFPAFWIYHLIYPLKGLMGRONRPTIYVWMSVLLA 1037
Db 598 VVAGSVDAINNGSGMPLFGKLFPAFWIYHLIYPLKGLMGRONRPTIYVWMSVLLA 657
QY 1038 IFSLLWVRIDPFTSRVTPGDIIECGI 1065
Db 658 IFSLLWVRIDPFTSRVTPGDIIECGI 685

RESULT 14

US-09-838-539-7
; Sequence 7, Application US/09838539
; Patent No. US20020129401A1
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; FILE REFERENCE: 15621/03/US
; CURRENT APPLICATION NUMBER: US/09/838,539
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; PRIOR APPLICATION NUMBER: 08/960,048
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 685
; TYPE: PRT
; ORGANISM: Gossypium hirsutum
US-09-838-539-7

Query Match 49.4%; Score 2803; DB 10; Length 685;
Best Local Similarity 74.5%; Pred. No. 1.5e-232;
Matches 512; Conservative 77; Mismatches 76; Indels 22; Gaps 6;

QY 396 ARKWPECKKYSIEPAPRPAKTDYLDKDYQTSFVADRDRAMKEFEFKIRIATLYS 455
Db 1 ARKWPECKKHNEPRAPPEFYENEKIDYLDKDYHPSFEVERRAMKEFEFKIRIATLYA 60
QY 456 KALKEPEGWVQDGPMPGNNTPGDMIOYFLGONGGLDAEGNELPRLVYVSREKRP 515
Db 61 KAKKREBEGWQDGPMPGNNTRDHPGMIQYFLGSAGALDVGKELPRLVYVSREKRP 120
QY 516 FOHHKAGAMNALIVRSVAVLTNGPFLINDCDHYINNSKALREAMCFMLDPNLGQVCY 575
Db 121 YOHKKAGAGAMNALIVRSVAVLTNAPFLINDCDHYINNSKALREAMCFMLDPNLGQVCY 180
QY 576 OFPQRFDDIDKDRANRNTVFEDINLGLDGIQGVYVYGGCVNFRKALYGEPRITV 635
Db 181 OFPQRFDDIDKDRANRNTVFEDINLGLDGIQGVYVYGGCVNFRKALYGEPRITV 240
QY 636 HKK-----PSLISKLKSGSKRSKRSKAKKE-----SDKKSGRH--TDSVTPVN 677
Db 241 RKMTCDCMPSMCCCCGSGSKRSKRSKKGKGLLGLLIGKKKKMMGKNYVKKSGAPVD 300
QY 678 LDDIEGVGAGFD--EKALLMSQMSLEKRFQSAVASTLMENGVPSPATPENILKE 736
Db 301 LDDIEGVGAGFD--EKALLMSQMSLEKRFQSAVASTLMENGVPSPATPENILKE 358
QY 737 AIHVISCGYEDKSDMGMEIGWITGSTEDILITGFKHARGMSIYMPKLPAGKAGAPIN 796
Db 796 AIHVISCGYEDKSDMGMEIGWITGSTEDILITGFKHARGMSIYMPKLPAGKAGAPIN 796

Db 359 AIHVISCGYEETKEMGEIGWITGSTEDILITGFKHARGMSVYCPKRPAGKAGAPIN 418
QY 797 LSRDLQVLRKMGALGSVEIFLSRHCPIMYGGKGLKMLERLAYINTIYPPFSIPLACYI 856
Db 419 LSRDLQVLRKMGALGSVEIFLSRHCPIMYGGKGLKMLERLAYINTIYPPFSIPLACYI 478
QY 857 LANCETNOFIIPQISNIASIWFLSLFISFATGILEMRSGVIGIDEMMRNEQFVIGY 916
Db 479 IPAVCLITGFIIPITLSNLTSVWFLALFSLIATGVLRLMSGSVIGIDEMMRNEQFVIGY 538
QY 917 VSAHLFAVFOGILKVLAGIDNTFTVTSKASDEGDAELYLKFWTLLIPPTLLIPLVNLV 976
Db 539 VSAHLFAVFOGILKVLAGIDNTFTVTSKASDEGDAELYLKFWTLLIPPTLLIPLVNLV 597
QY 977 GVVAGSVYAINSGYSGMPLFGKLFPAFWIYHLIYPLKGLMGRONRPTIYVWMSVLLA 1036
Db 598 GVVAGSVDAINNGSGMPLFGKLFPAFWIYHLIYPLKGLMGRONRPTIYVWMSVLLA 657
QY 1037 IFSLLWVRIDPFTSRVTPGDIIECGI 1063
Db 658 IFSLLWVRIDPFTSRVTPGDIIECGI 684

RESULT 15

US-09-900-237-16
; Sequence 16, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen
; TITLE OF INVENTION: Plant Cellulose Synthases
; FILE REFERENCE: B81170 US CIP
; CURRENT APPLICATION NUMBER: US/09/900,237
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: 60/092,844
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: PCT/US99/15871
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: 09/720383
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Glycine max
US-09-900-237-16

Query Match 46.6%; Score 2647.5; DB 10; Length 610;
Best Local Similarity 78.8%; Pred. No. 3.1e-219;
Matches 482; Conservative 55; Mismatches 72; Indels 3; Gaps 2;

QY 455 SKALKEPEGWVQDGPMPGNNTPGDMIOYFLGONGGLDAEGNELPRLVYVSREKRP 514
Db 1 AKAKPPEBEGWVQDGPMPGNNTRDHPGMIQYFLGSHSGDLTDGELPRLVYVSREKRP 60
QY 515 GFOHHKAGAMNALIVRSVAVLTNGPFLINDCDHYINNSKALREAMCFMLDPNLGQVCY 574
Db 61 GFOHHKAGAMNALIVRSVAVLTNGPFLINDCDHYINNSKALREAMCFMLDPNLGQVCY 120
QY 575 VOFQRFDDIDKDRANRNTVFEDINLGLDGIQGVYVYGGCVNFRKALYGEPRITV 634
Db 121 VOFQRFDDIDKDRANRNTVFEDINLGLDGIQGVYVYGGCVNFRKALYGEPRITV 180
QY 635 HKKPSLISKLKSGSKRSKRSKAKKE-----SDKKSGRH--TDSVTPVN 677
Db 181 EDIEPVITVSCCGSKRSKRSKKGKGLLGLLIGKKKKMMGKNYVKKSGAPVD 300
QY 694 KALLMSQMSLEKRFQSAVASTLMENGVPSPATPENILKEAIVHISGYEDKSDMG 753
Db 239 RLLMSQMSLEKRFQSAVASTLMENGVPSPATPENILKEAIVHISGYEDKSDMG 298
QY 754 EIGWITGSTEDILITGFKHARGMSIYMPKLPAGKAGAPINSLRLQVLRKMGALGSVE 813
Db 813 EIGWITGSTEDILITGFKHARGMSIYMPKLPAGKAGAPINSLRLQVLRKMGALGSVE 813

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 16, 2003, 11:05:53 ; Search time 18 Seconds
(without alignments)
1740.856 Million cell updates/sec

Title: AAC39336
Perfect score: 5677
Sequence: 1 MESEGETAGKPMKNIVPQC.....IDPFSRYTGPDLECGINC 1065

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/Dackfilest.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4133.5	72.8	881	US-08-960-048-8	Sequence 8, Appl
2	3473.5	61.2	974	US-08-960-048-6	Sequence 6, Appl
3	2803	49.4	685	US-08-960-048-7	Sequence 7, Appl
4	278	4.9	693	US-08-960-048-11	Sequence 11, Appl
5	245.5	4.3	861	US-08-960-048-12	Sequence 12, Appl
6	236.5	4.2	756	US-08-960-048-10	Sequence 10, Appl
7	228.5	4.0	756	US-09-147-236-3	Sequence 3, Appl
8	228.5	4.0	3031	US-07-689-008-2	Sequence 2, Appl
9	196	3.5	723	US-08-960-048-9	Sequence 9, Appl
10	196	3.5	723	PCT-US91-01726-4	Sequence 4, Appl
11	124	2.2	357	US-08-119-773-4	Sequence 4, Appl
12	116	2.0	357	US-08-119-773-6	Sequence 6, Appl
13	115	2.0	346	US-08-119-773-5	Sequence 5, Appl
14	115	2.0	357	US-08-119-773-2	Sequence 2, Appl
15	110	1.9	1587	US-09-000-094-46	Sequence 46, Appl
16	107	1.9	686	US-08-768-301-4	Sequence 4, Appl
17	106.5	1.9	1437	US-09-061-400-2	Sequence 2, Appl
18	106.5	1.9	1453	US-09-001-273-2	Sequence 2, Appl
19	106.5	1.9	1453	US-08-843-459A-2	Sequence 2, Appl
20	104	1.8	1167	US-08-485-568A-6	Sequence 6, Appl
21	104	1.8	1167	US-08-590-554A-6	Sequence 6, Appl
22	104	1.8	1167	US-09-184-223-6	Sequence 6, Appl
23	104	1.8	3033	US-07-925-695-5	Sequence 5, Appl
24	103	1.8	1168	US-08-620-717A-9	Sequence 9, Appl
25	102.5	1.8	1871	US-08-694-869-1	Sequence 1, Appl
26	102.5	1.8	1871	US-09-349-546-1	Sequence 1, Appl
27	101.5	1.8	765	US-08-425-061-19	Sequence 19, Appl

28	101.5	1.8	765	US-08-825-886-19	Sequence 19, Appl
29	101.5	1.8	900	US-08-425-061-20	Sequence 20, Appl
30	101.5	1.8	900	US-08-825-886-20	Sequence 20, Appl
31	101.5	1.8	914	US-08-425-061-21	Sequence 21, Appl
32	101.5	1.8	914	US-08-825-886-21	Sequence 21, Appl
33	101.5	1.8	1202	US-08-425-061-22	Sequence 22, Appl
34	101.5	1.8	1202	US-08-825-886-22	Sequence 22, Appl
35	101.5	1.8	1363	US-08-425-061-23	Sequence 23, Appl
36	101.5	1.8	1363	US-08-825-886-23	Sequence 23, Appl
37	101.5	1.8	1852	US-08-425-061-24	Sequence 24, Appl
38	101.5	1.8	1852	US-08-825-886-24	Sequence 24, Appl
39	101.5	1.8	1863	US-08-598-591-2	Sequence 2, Appl
40	101.5	1.8	1863	US-08-480-784-2	Sequence 2, Appl
41	101.5	1.8	1863	US-08-480-784-2	Sequence 2, Appl
42	101.5	1.8	1863	US-08-487-003-2	Sequence 2, Appl
43	101.5	1.8	1863	US-08-483-554B-2	Sequence 2, Appl
44	101.5	1.8	1863	US-08-798-691-2	Sequence 2, Appl
45	101.5	1.8	1863	US-08-798-691-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1					
US-08-960-048-8					
Sequence 8, Application US/08960048C					
Patent No. 6271443					
GENERAL INFORMATION:					
APPLICANT: Stalker, D. et al.					
TITLE OF INVENTION: Plant Cellulose Synthase and Promoter					
FILE REFERENCE: 15621/01/US					
CURRENT APPLICATION NUMBER: US/08/960,048C					
CURRENT FILING DATE: 1997-10-29					
PRIOR APPLICATION NUMBER: 60/029,987					
PRIOR FILING DATE: 1996-10-29					
NUMBER OF SEQ ID NOS: 12					
SOFTWARE: FastSeq for Windows Version 4.0					
SEQ ID NO 8					
LENGTH: 881					
TYPE: PRT					
ORGANISM: Oryzae sativa					
US-08-960-048-8					
Query Match					
Best Local Similarity 72.8%; Score 4133.5; DB 4; Length 881;					
Matches 766; Conservative 55; Mismatches 56; Indels 5; Gaps 3;					
QY	188	GNVAMKERYDGMKMKOEKNTGPNV---SNOAASERGV--DIDASTDILADEALLNDEARHP	243		
DB	1	GNVAMKERYDGMKMKOEKNTGPNV---SNOAASERGV--DIDASTDILADEALLNDEARHP	60		
QY	244	LSRKVSIPSSRIINRYIMLVLCFLHYRITPNVPAFALVSVICETWALSMT	303		
DB	61	LSRKVPDSSRIINRYIMLVLCFLHYRITPNVPAFALVSVICETWALSMT	120		
QY	304	LDQFKKRPVARETFLDLRLALRYDREGEPSQAAVDIRYSTYDPLKEPLVYANVLSL	363		
DB	121	IDQFKKRPVARETFLDLRLALRYDREGEPSQAAVDIRYSTYDPLKEPLVYANVLSL	180		
QY	364	AVDYVVDVSCVYFDGAAAMLSFESLAETSEFARKMVPFCKYSLEPRAPERYFAKIDY	423		
DB	181	AVDYVVDVSCVYFDGAAAMLSFESLAETSEFARKMVPFCKYSLEPRAPERYFAKIDY	240		
QY	424	LKDQVOTSEVXDRAMKREVEEFKTRINALVSKALCKCEGVMODGTPWPNNTGDHPG	483		
DB	241	LKDQVOTSEVXDRAMKREVEEFKTRINALVSKALCKCEGVMODGTPWPNNTGDHPG	300		
QY	484	MIQVFLGONGCIGDAGNELPRLYVYSREKRPFOHKKAGAMNALVRSAYLTNGPFLN	543		
DB	301	MIQVFLGONGCIGDAGNELPRLYVYSREKRPFOHKKAGAMNALVRSAYLTNGPFLN	360		
QY	544	LDGDHYINNSKALREAMCEFLMDPNNGKQVYQFQDRDGDIDKNDRYANRNTVFVDILNR	603		


```

Db      361 LDCDHINNSKALREMAFELMDPNLGRSVYQFOPRFDGIDRNDRYANNNTVFEDINR 420
OY      604 GIDGIGPYVYGCGEJFNRTALYGEPPYKVKHKKRSLISLKGGRKNSKAKKSDKK 663
Db      421 GIDGIGPYVYGCGEJFNRTALYGEPPYKVKHKKRSLISLKGGRKNSKAKKSDKK 479
OY      664 KSGRHTDSTVPVFNLDIEGEGAGFDEDEKALIMSOMSLKFRGSAVFAVSTLMENG 723
Db      480 KSNHVDASAVPFNEDIEGEGAGFDEDEKALIMSOMSLKFRGSAVFAVSTLMENG 539
OY      724 VPPSAPFENLKEAHIVISCGYEDKSDMGMEIGMITYGTEDILTFKMHARGMSIYCM 783
Db      540 VQOSAPFENLKEAHIVISCGYEDKSDMGMEIGMITYGTEDILTFKMHARGMSIYCM 599
OY      784 PLPAKGSAPINLSRNLNOVLKMAVSEILSRHCPIMYGNGRLKFLERAYVNTI 843
Db      600 PKRPAKGSAPINLSRNLNOVLKMAVSEILSRHCPIMYGNGRLKFLERAYVNTI 659
OY      844 YPITSIPILMYCTLLAVCLFTNOFIPOISINIASIMFLSLFISFATGILEMRSGVGD 903
Db      660 YPITSIPILMYCTLLAVCLFTNOFIPOISINIASIMFLSLFISFATGILEMRSGVGD 719
OY      904 EWMRNEQFVIGVSAHLFAVFOGILKVLADIDNTFTYKSKASDEGDEPAELYLFKWTTL 963
Db      720 EWMRNEQFVIGVSAHLFAVFOGILKVLADIDNTFTYKSKASDEGDEPAELYLFKWTTL 779
OY      964 LIPPTLLILNVGVAVGSAVAINSGYOMPLGCLFFAFVYVHLVLPFLGLMGRONR 1023
Db      780 LIPPTLLILNVGVAVGSAVAINSGYOMPLGCLFFAFVYVHLVLPFLGLMGRONR 839
OY      1024 PPTIVVWVLLASIFSLMVRIDPFTSRVGTGPDILECGINC 1065
Db      840 PPTIVVWVLLASIFSLMVRIDPFTSRVGTGPDILECGINC 881

```

```

RESULT 2
US-08-960-048-6
; Sequence 6, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; FILE REFERENCE: 15621/01/US
; CURRENT APPLICATION NUMBER: US/08/960,048C
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/029,987
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 974
; TYPE: PRT
; ORGANISM: Gossypim hirsutum
US-08-960-048-6

```

```

Query Match      61.2%; Score 3473.5; DB 4; Length 974;
Best Local Similarity 62.1%; Pred. No. 0;
Matches 660; Conservative 129; Mismatches 176; Indels 97; Gaps 15;

```

```

OY      12 MNIVPQTQICSDNVKTVGDRAVACDICSFPVRCPEYERKDGNSCPQCKTRRKR 71
Db      2 MESSGPV-VCHTCEHVLNVGEBPFAVACHCNPIKCSFEYDLKGRACLRG----- 54
OY      72 LKGSRAIPGDKDEGLADGTYEENPQKEKISERLGMHLGRGKEEMGEPOYDKREVSH 131
Db      55 --GSP-----YDENLDD-----VERATGDOSTMAHL--NKSQDVG-----IHA 90
OY      133 NMLPRLTSRODTSGEFSASPERLSVSTIAGKRLPYSSDVNQSNNRIYDPVGLGNVA 191
Db      91 RHISVSSTIDSEMAEDN-----GNSI 111

```

```

OY      192 WKERVDMKMKQKCKNTGPVSTQASERGVGDIDASTDLADEALLND---EAROPLSRK 247
Db      112 MKRVSWMKREKKKKKKPATTKEVERAE-----IPPEQOMEDKKPAPDASQPLSTI 161
OY      248 VSIPISSRINPYRANVIMLRVILCLFLHYATITNPVPAFAFLMNLVSVICELTFALSWLDOF 307
Db      162 IPIPKSRAPYRFRVITIMRLILGLFPHYVNTNPVDASAFGLMILTYSVCELMFAFSWLDQF 221
OY      308 PKMFPVNRREYLDRLARLDREREQESQLAVIDLFVSTVPLKEPPLVNTANTVLSILAVY 367
Db      222 PKMFPVNRREYLDRLARLDREREQESQLAVIDLFVSTVPLKEPPLVNTANTVLSILAVY 281
OY      368 PVDKVCYVFDGGAAMLSESLAETSEFARKVVPFCCKYSIEPRADWYFAAKIDYLDK 427
Db      282 PVDKVCYVFDGGAAMLSESLAETSEFARKVVPFCCKYSIEPRADWYFAAKIDYLDK 341
OY      428 VQTSFVKDRRAKREIEEFKIRINALVSAKLCREGEVWMOGTPMPRGWNTGDHPGMIOY 487
Db      342 VQTSFVKDRRAKREIEEFKIRINALVSAKLCREGEVWMOGTPMPRGWNTGDHPGMIOY 401
OY      488 FLGONGGLDAEGNELPRLVYVSREKRPGRQHHKAGAMNALVSAVLNNGPFIILNLD 547
Db      402 FLGONGGLDAEGNELPRLVYVSREKRPGRQHHKAGAMNALVSAVLNNGPFIILNLD 461
OY      548 HYINNSKALREACFLMDPNLGRSVYQFOPRFDGIDRNDRYANNNTVFEDINR 607
Db      462 HYINNSKALREACFLMDPNLGRSVYQFOPRFDGIDRNDRYANNNTVFEDINR 521
OY      608 IGPVYVGTGCVFNRALYGEPPYKVKHKKRSLISLKGGRKNSKAKKSDKK 667
Db      522 IGPVYVGTGCVFNRALYGEPPYKVKHKKRSLISLKGGRKNSKAKKSDKK 579
OY      668 HDSSTVPVFNLDIEGEGAGFDD--EKALIMSOMSLKFRGSAVFAVSTLMENG 726
Db      580 HDSSTVPVFNLDIEGEGAGFDD--EKALIMSOMSLKFRGSAVFAVSTLMENG 633
OY      727 SATPENLKEAHIVISCGYEDKSDMGMEIGMITYGTEDILTFKMHARGMSIYCM 786
Db      634 SATPENLKEAHIVISCGYEDKSDMGMEIGMITYGTEDILTFKMHARGMSIYCM 693
OY      787 PAFKGSAPINLSRNLNOVLKMAVSEILSRHCPIMYGNGRLKFLERAYVNTI 845
Db      694 PAFKGSAPINLSRNLNOVLKMAVSEILSRHCPIMYGNGRLKFLERAYVNTI 753
OY      846 ITSIPILMYCTLLAVCLFTNOFIPOISINIASIMFLSLFISFATGILEMRSGVGD 905
Db      754 ITSIPILMYCTLLAVCLFTNOFIPOISINIASIMFLSLFISFATGILEMRSGVGD 813
OY      906 WRNEQFVIGVSAHLFAVFOGILKVLADIDNTFTYKSKASDEGDEPAELYLFKWTTL 965
Db      814 WRNEQFVIGVSAHLFAVFOGILKVLADIDNTFTYKSKASDEGDEPAELYLFKWTTL 872
OY      966 PPTLLILNVGVAVGSAVAINSGYOMPLGCLFFAFVYVHLVLPFLGLMGRONR 1025
Db      873 PPTLLILNVGVAVGSAVAINSGYOMPLGCLFFAFVYVHLVLPFLGLMGRONR 932
OY      1026 PPTIVVWVLLASIFSLMVRIDPFTSRVGTGPDILECGINC 1065
Db      933 PPTIVVWVLLASIFSLMVRIDPFTSRVGTGPDILECGINC 974

```

```

RESULT 3
US-08-960-048-7
; Sequence 7, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; FILE REFERENCE: 15621/01/US
; CURRENT APPLICATION NUMBER: US/08/960,048C
; PRIOR FILING DATE: 1997-10-29

```


Db 453 ASKINKMR-YSEFSEVETMALFLVRYTVITLLSPRGKENVDPKGGLEKGYFDL- 509
Qy 958 FKMTLLIPTPTLLIVNLVGVAGSYAINSGYOSWGLGRLFFAFVYVYHLFPFLKGL 1017
Db 510 ---GAVYPNIIILGIMFGLARGV-YELSPGHLDQIAERAYLLNSAWAMLSLIIILAI 564
Qy 1018 -MGRO-----NRTPTIVV 1030
Db 565 AVGRETOOKRNSHRIPATIPV 585

RESULT 10
PCT-US91-01726-4
Sequence 4, Application PC/TUS9101726
GENERAL INFORMATION:
APPLICANT: Brown, Malcolm R
APPLICANT: Saxena, Indar M
APPLICANT: Lin, Fong C
TITLE OF INVENTION: Recombinant Cellulose Synthase
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: David L. Parker
STREET: 600 Congress Avenue Suite 2300
CITY: Austin
STATE: Texas
COUNTRY: United States
ZIP: 78701

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/01726
FILING DATE: 19910314
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 494093
FILING DATE: 15-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Parker Mr, David L
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512-320-7200
TELEFAX: 512-474-7577
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 723 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
PCT-US91-01726-4

Query Match 3.5%; Score 196; DB 5; Length 723;
Best Local Similarity 20.1%; Pred. No. 1.9e-10;
Matches 161; Conservative 93; Mismatches 241; Indels 306; Gaps 36;

Qy 251 PSSRINPRMYIMLFLVLCFLHAYRITNPVNAFALM-----VSYICEIWEPLSWIL 304
Db 70 PSSRSQIF-LEVLSSIVSL-RYLWRLTETL--SFDTWLOGLLGMLLVAELIYALMMLFL 125
Qy 305 DQFPMFVFNRETYIDRLALRYDRGCEPSQLAANDIFVSTVDPLKEPPLVTANTVLSILA 364
Db 126 SYFQIADPLHRP-----LPL-----PPNDEWPTVDIFVPTYN---EELISVRLVLSIG 174
Qy 365 VDYPVADKVCYVDDGAAMLSPESIAETSEFARKWVPCKKYSIEPRAPEWYFAKIDYL 424
Db 175 IDMPREKRVHILDDG-----RRPE--FAA----- 197
Qy 425 KDKVOTSYVKDRBRAMKREYEERKINALVSKALKCPREGVWMDGTPMPGNNGTDDHPCM 484
Db 198 -----FAAEC----- 202

Qy 485 IOVFLGONGGLDAEGNEELPRLYVYSRERKPGFOHHKKAGAMNALRVSAVLNCPPIIL 544
Db 203 ---GAN-----YIARPT-----NEAAKGNLNYAIG---HTDGYILLF 235
Qy 545 DCDHYINNSKALREAMCFILM-DPNLKOVCYVQPORREDGIDKNDRYANRNTVFEDINLR 603
Db 236 DCDH-VPTRAFLQITMGWVDEDP-----KIALMQTPHHFSPDP----- 273
Qy 604 GLDGIQGVYVVGCGVFNRTALXGEPRPIKVKHKKRPSLSKICGSGRRKNSAKKESDOK 663
Db 274 -----FORNLSAGYRTPT----- 285
Qy 664 KSGRHTDSTVPVFNIDIDEEGVGAGFDEKALLMSOHSLEKRFQSAVFVASTLMEENG 723
Db 286 -----PEGNL---FYGVYQDGN-----FNDATFFGSCAI----- 313
Qy 724 VPPSATPENLKEAIVHISCGYEDKSDMGEIGWISYTEDILTGFKHARGMSIYCM 783
Db 314 -----LRTALEQIG-----GFATQVTEBAHTALMKQRLGWSYAYL- 350
Qy 784 KRLPAKGSAPRLNLSDRNLQVLRNALGSVEILFSRHCPTWYGNR-LKFLERFAYVNT 842
Db 351 -RIPLAGGLATERLLHIGQVRMARGMLOI-FRIDNPLE---GRGLSWGQRLCYLSAM 404
Qy 843 IYPIRSIPLMYCTLLAVCLFTNPFILPOISNIASIMFLSLPSIFATGILMRSGVG- 901
Db 405 TSFLFAVRVIFLSSPLAFLEFGQNIITASP-----LALLAIAIPM-FHAYGT 452
Qy 902 ---IDEMWRNEQFW-VIGVSAHLFAVRGILKVLAGIDTNETVTSKASDEGDPAEVL 957
Db 453 ASKINKMR-YSEFSEVETMALFLVRYTVITLLSPRGKENVDPKGGLEKGYFDL- 509
Qy 958 FKMTLLIPTPTLLIVNLVGVAGSYAINSGYOSWGLGRLFFAFVYVYHLFPFLKGL 1017
Db 510 ---GAVYPNIIILGIMFGLARGV-YELSPGHLDQIAERAYLLNSAWAMLSLIIILAI 564
Qy 1018 -MGRO-----NRTPTIVV 1030
Db 565 AVGRETOOKRNSHRIPATIPV 585

RESULT 11
US-08-119-773-4
Sequence 4, Application US/08119773
Patent No. 5460942
GENERAL INFORMATION:
APPLICANT: Chou, Janice Y.
APPLICANT: Lei, Ke-Jian
APPLICANT: Shelly, Leslie L.
TITLE OF INVENTION: GLUCOSE-6-PHOSPHATASE: THE GENE AND
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourile and Crew
STREET: Stewart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/119,773
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-175
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 357 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1..357
OTHER INFORMATION: /note= "The G-6-Pase amino acid sequence mutated
OTHER INFORMATION: from Arg to Cys at position 83."
US-08-119-773-4

Query Match 2.2%; Score 124; DB 1; Length 357;
Best Local Similarity 19.8%; Pred. No. 0.0011;
Matches 79; Conservative 44; Mismatches 122; Indels 154; Gaps 16;

QY 736 EAIHVISCYEDKSDGMEIGWYGSVTEIDLTGFKMARGRSICYMKLPKFGSAPI 795
DB 14 GSTHYLVQVYQDSQD-----FLVSVIADLRNAYVLPPIW---FHLQEAAGIKLLMVA 65
QY 796 NISDRNLNOLRMAALSGVELLFSRHCPHWG----- 825
DB 66 VIGDWLNLVEKW-----LTFGQ-CPYWVLDTDYYSNTSVPLIKOPVTCETGSGPSG 118
QY 826 -----YNGRLKFLERFAYVNTTIYPTSIPILMYCTLLAVCL----- 862
DB 119 HAMGTAGYVYVAVTSTLSIFQCKIKPTYFRCLNVILM-----LGFMAVOLNVCISRTY 172
QY 863 ----FTNOFTIPOISNIA-----SIWFLSL-----FLSIFATGILEMRSGVG 901
DB 173 LAHPHQVAVAGVLSGIATVETFSHISYNAKLKYPFLTFPFSFAIGFY-LLLKIGG 231
QY 902 IDEWNRNE--QFWVIGVSAHLFAVQGLKVLAGIDTFYTSKASDEDDGFAELYLEK 959
DB 232 VDLMTLEKAQWCEQPEKWH-----IDTT----- 256
QY 960 WTTLLIPTTLIIIVLVGAVGVSAINS-----YOSGCLFGKLPFAFWYVHL 1010
DB 257 -----PRASLLKNL-GTLFGLGLALNSMTRESCKGLSKWLPFLCSIVASIVLLHV 308
QY 1011 YPFLKGLMGRNRPPIVYVMSVLASIFSLMVRIDPF 1049
DB 309 FDSLK---PPSQVELVEFVLSFCKSAVAVPLASVSIVPY 343

RESULT 12
US-08-119-773-6

; Sequence 6, Application US/08119773
; Patent No. 5460942
; GENERAL INFORMATION:
; APPLICANT: Chou, Janice Y.
; APPLICANT: Lei, Ke-Jian
; APPLICANT: Shelly, Leslie L.
; TITLE OF INVENTION: GLUCOSE-6-PHOSPHATASE: THE GENE AND
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: Stewart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105-1493
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/119,773
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-175
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 357 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1..357
OTHER INFORMATION: /note= "G-6-Pase mutated at position 295 with Arg
OTHER INFORMATION: changed to Cys"
US-08-119-773-6

Query Match 2.0%; Score 116; DB 1; Length 357;
Best Local Similarity 19.5%; Pred. No. 0.0073;
Matches 78; Conservative 44; Mismatches 123; Indels 154; Gaps 16;

QY 736 EAIHVISCYEDKSDGMEIGWYGSVTEIDLTGFKMARGRSICYMKLPKFGSAPI 795
DB 14 GSTHYLVQVYQDSQD-----FLVSVIADLRNAYVLPPIW---FHLQEAAGIKLLMVA 65
QY 796 NISDRNLNOLRMAALSGVELLFSRHCPHWG----- 825
DB 66 VIGDWLNLVEKW-----LTFGQ-CPYWVLDTDYYSNTSVPLIKOPVTCETGSGPSG 118
QY 826 -----YNGRLKFLERFAYVNTTIYPTSIPILMYCTLLAVCL----- 862
DB 119 HAMGTAGYVYVAVTSTLSIFQCKIKPTYFRCLNVILM-----LGFMAVOLNVCISRTY 172
QY 863 ----FTNOFTIPOISNIA-----SIWFLSL-----FLSIFATGILEMRSGVG 901
DB 173 LAHPHQVAVAGVLSGIATVETFSHISYNAKLKYPFLTFPFSFAIGFY-LLLKIGG 231
QY 902 IDEWNRNE--QFWVIGVSAHLFAVQGLKVLAGIDTFYTSKASDEDDGFAELYLEK 959
DB 232 VDLMTLEKAQWCEQPEKWH-----IDTT----- 256
QY 960 WTTLLIPTTLIIIVLVGAVGVSAINS-----YOSGCLFGKLPFAFWYVHL 1010
DB 257 -----PRASLLKNL-GTLFGLGLALNSMTRESCKGLSKWLPFLCSIVASIVLLHV 308
QY 1011 YPFLKGLMGRNRPPIVYVMSVLASIFSLMVRIDPF 1049
DB 309 FDSLK---PPSQVELVEFVLSFCKSAVAVPLASVSIVPY 343

RESULT 13
US-08-119-773-5

; Sequence 5, Application US/08119773
; Patent No. 5460942
; GENERAL INFORMATION:
; APPLICANT: Chou, Janice Y.
; APPLICANT: Lei, Ke-Jian
; APPLICANT: Shelly, Leslie L.

TITLE OF INVENTION: GLUCOSE-6-PHOSPHATASE: THE GENE AND
TITLE OF INVENTION: PROTEIN AND RELATED MUTATIONS
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourie and Crew
STREET: Steuart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/119,773
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-175
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 346 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1..346
OTHER INFORMATION: /label= Truncated prot
OTHER INFORMATION: /note= *G-6-pase truncated due to a C to T
OTHER INFORMATION: mutation at nucleotide 1118 converting a glutamine
OTHER INFORMATION: to a stop codon at 347*
US-08-119-773-5

Query Match 2.0%; Score 115; DB 1; Length 346;
Best Local Similarity 19.5%; Pred. No. 0.0087;
Matches 78; Conservative 44; Mismatches 123; Indels 154; Gaps 16;
QY 736 EAIHVISCYEDKSDMGMEIGYGVYEDILITGFKHARGRSIYCPKLPAPFGSAPI 795
DB 14 OSTRHLYQVYDSDQM-----FILSVIADLRNAPYVLPPIW--FHLQEA VGIKLLNVA 65
QY 796 NLSDRNLQVLRWALGSVEILFSRHCPYWG----- 825
DB 66 VIGDWLNLVFKW-----ILFGQR-PYMWVLDTDYVNTSVPLIKOPVTCETGPGSPG 118
QY 826 -----YNGRLKFLERFAVNTTIPITSIPLMCTLLAVCL----- 862
DB 119 HAMGTAGYVYVWVTSLISIFQGIKIPYRFRCLNVIW-----LGFNAVOLNCLSLRIY 172
QY 863 -----FTNOFIIPQISNIA-----SIFLSL-----FLSIFATGILEMRWSGVG 901
DB 173 LAHFPHQVAVGVSGIAVTETFSIHSHIYNASLKKYFLITFFLFSFAIGFY-LLLKGLG 231
QY 902 IDEWMRNE--QFWVIGVSAHLFAVFOGILKVLAGIDTNFTVTSKASDEDDGFALYLFK 959
DB 232 VDLMTLEKQRMCEQPEMVH-----IDYT----- 256
QY 960 WTTLLIPPTLLIYNLVGVAVGSYAINSG-----YOSWGPLFGKLPFAFWYVYHL 1010
DB 257 -----PFASLKLNLTGLFGLIALNMSWYRESCKGLSKWLPFRLLSIYASIVLVLHV 308

QY 1011 YPEKLGMRONRPTIYVWWSVLLASIFSLLWRIDPF 1049
DB 309 FDSLK-----PPSQVELVIFYVLSFCKSAVYPLASVIRY 343
RESULT 14
US-08-119-773-2
Sequence 2, Application US/08119773
Patent No. 5460942
GENERAL INFORMATION:
APPLICANT: Chou, Janice Y.
APPLICANT: Lei, Ke-Jian
APPLICANT: Shelly, Leslie L.
TITLE OF INVENTION: GLUCOSE-6-PHOSPHATASE: THE GENE AND
TITLE OF INVENTION: PROTEIN AND RELATED MUTATIONS
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourie and Crew
STREET: Steuart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/119,773
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-175
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-5043
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 357 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-119-773-2
Query Match 2.0%; Score 115; DB 1; Length 357;
Best Local Similarity 19.5%; Pred. No. 0.0092;
Matches 78; Conservative 44; Mismatches 123; Indels 154; Gaps 16;
QY 736 EAIHVISCYEDKSDMGMEIGYGVYEDILITGFKHARGRSIYCPKLPAPFGSAPI 795
DB 14 OSTRHLYQVYDSDQM-----FILSVIADLRNAPYVLPPIW--FHLQEA VGIKLLNVA 65
QY 796 NLSDRNLQVLRWALGSVEILFSRHCPYWG----- 825
DB 66 VIGDWLNLVFKW-----ILFGQR-PYMWVLDTDYVNTSVPLIKOPVTCETGPGSPG 118
QY 826 -----YNGRLKFLERFAVNTTIPITSIPLMCTLLAVCL----- 862
DB 119 HAMGTAGYVYVWVTSLISIFQGIKIPYRFRCLNVIW-----LGFNAVOLNCLSLRIY 172
QY 863 -----FTNOFIIPQISNIA-----SIFLSL-----FLSIFATGILEMRWSGVG 901
DB 173 LAHFPHQVAVGVSGIAVTETFSIHSHIYNASLKKYFLITFFLFSFAIGFY-LLLKGLG 231
QY 902 IDEWMRNE--QFWVIGVSAHLFAVFOGILKVLAGIDTNFTVTSKASDEDDGFALYLFK 959
DB 232 VDLMTLEKQRMCEQPEMVH-----IDYT----- 256
QY 960 WTTLLIPPTLLIYNLVGVAVGSYAINSG-----YOSWGPLFGKLPFAFWYVYHL 1010
DB 257 -----PFASLKLNLTGLFGLIALNMSWYRESCKGLSKWLPFRLLSIYASIVLVLHV 308

Db 257 -----PRASLIKLN-GTLFGIGLALNNSMYRESCGKLSKWLPERLSIVASVLLHV 308
QY 1011 YPLKGLMGRONRTPTIVVWSVLASISFSLMWRIIDPF 1049
Db 309 FDSLK-----PPSQVELVFLYLSFKCSAVPLASVSIPIY 343

RESULT 15
US-09-000-094-46
; Sequence 46, Application US/09000094
; Patent No. 6365160

GENERAL INFORMATION:
APPLICANT: WEBB, Elizabeth Ann
MARGETTS, Mary Bridgid
COX, John Cooper
FRAZER, Ian
MCMILLAN, Nigel Alan John
WILLIAMS, Mark Philip
MOLONEY, Margaret Bridget
Holland
EDWARDS, Stirling John
TITLE OF INVENTION: PAPILLOMAVIRUS POLYPEPTIDE CONSTRUCTS
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/09/000,094
FILING DATE: 21-Apr-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/AU96/00473
FILING DATE: 26-JUL-1996
APPLICATION NUMBER: AU PN 4439/95
FILING DATE: 27-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 017227/0137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5399
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1587 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-000-094-46

Query Match 1.98; Score 110; DB 4; Length 1587;
Best Local Similarity 19.38; Pred. No. 0.44;
Matches 114; Conservative 77; Mismatches 221; Indels 180; Gaps 30;

QY 2 ESEGETAGKPMK---NIVPOTQICSDNDGKTYDGRFVACDICSFPVCRPCVEYERKDC 58
Db 877 EVDGDS-OPLKQHFQIVTCCG-CDSNVRLVYQ-----CTETDIREV 917
QY 59 NQ-----SCPOCKTRYKRLKGSIPAIGDKDEGLADEGVTFVFNYPQEKISERMGLM 110
Db 918 QOLLGLTINIVPIC-----APKTSMDDDSTENEGSCTGMFVNAIVQHPGTGT 968
QY 111 HLTRGGEEMGEPOY-----DKEVSHNHL--PRLTSRODTSGEFSASPERLSVSSSTI 161

Db 969 QISDDEDEVEDSGYDMVDFIDDSNITHNSLEAQLFNQEAOTHY-----ATY 1017
QY 162 AGGKR-----LPYSSDNO-----SPRRRTVDVVGGNVAMK-----ERV-D-G 198
Db 1018 QDLKRRYLSPPVSPINTIAEAVSEISPR--LDAIKLTPQCKVKKRLLFOTRELDSG 1074
QY 199 WKMKQ-EKNTG-PVSTQAAASERG--GVDDIDASTDIADBALNDEARQPLSRKVSIPSSR 254
Db 1075 YGSEVEAGTGTQVEKHGVPENGDGOQENDTGRDIEGE-----HTEAEAPTNS 1123
QY 255 INPYRMVIMRLVILCLFLHYRITNPVNAFALMYSVI-----CEINFALSM--- 302
Db 1124 VREHAGTAGLIELLKCKDLRAALLGFKKBCFGISFLDIRPKSDKTTCLDWVVAQFGIH 1183
QY 303 --LLDQPKWF-PVNBETYLDRLA-----LRYDREGERSOLA---AVDIFYSTVD 346
Db 1184 HSISEAFQKLIPELUSYAHIQWLTNMGVLLVLFKPKVKSSTVAFRLATLTLIPENO 1243
QY 347 PLKEPPLV-----TANTVLSILAVDYVDKVSQVFPDDGAAMLSPF----- 387
Db 1244 MLEPPKIQSGVALYWPFTGISNASTVIG---EAPENITRTQVTEHGLAQSQFLTEH 1299
QY 388 -----SLAETSEPRARKVPPCKKYSIEPRAPETFAAKIDYLRKQVOTSFYKDRRAK 440
Db 1300 VQWAYDNDICEESEIHA---FEYAQRGDPDFSNARA-----FLNSNMQAKYVKDCATMC 1348
QY 441 REYEERKIRTNALVSKALCPREGVWMOGTTPAGNNTGDHGMIOVFIQON 492
Db 1349 RHYKHAEMR-----KMSIKQWIKHRSKITEG--TQNMKPIVOFLRHON 1389

Search completed: June 16, 2003, 11:08:57
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:01:28 ; Search time 45 Seconds
(without alignments)
3153.598 Million cell updates/sec

Title: AAC39336
Perfect score: 5677
Sequence: 1 MESEGETAGKPMKNIVPQTC.....IDPFTSRVTGPDIIECGINC 1065

Scoring table:
BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.GeneSeq_101002:*

- 1: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:*
- 2: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:*
- 3: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:*
- 4: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:*
- 5: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:*
- 6: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:*
- 7: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:*
- 8: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:*
- 9: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:*
- 10: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:*
- 11: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:*
- 12: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:*
- 13: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:*
- 14: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:*
- 15: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:*
- 16: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:*
- 17: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:*
- 18: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:*
- 19: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:*
- 20: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:*
- 21: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:*
- 22: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:*
- 23: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5655	99.6	1065	19	AAW3819
2	5654	99.6	1065	21	AAW3819
3	5654	99.6	1065	23	AAW3819
4	5597	98.6	1054	21	AAW3819
5	5066	89.2	959	21	AAW3819
6	4572	80.5	1077	21	AAW3819
7	4572	80.5	1077	21	AAW3819
8	4564	80.4	1079	23	AAW3819
9	4529.5	79.8	1076	21	AAW3819
10	4418	77.8	1043	21	AAW3819

11	4028.5	71.0	1075	21	AAW3819
12	4028.5	71.0	1075	21	AAW3819
13	4028.5	71.0	1075	21	AAW3819
14	4024	70.9	1074	21	AAW3819
15	4024	70.9	1074	21	AAW3819
16	4024	70.9	1074	21	AAW3819
17	3984	70.2	1081	19	AAW3819
18	3984	70.2	1081	19	AAW3819
19	3974	70.0	1081	19	AAW3819
20	3974	70.0	1081	19	AAW3819
21	3900	68.7	1058	21	AAW3819
22	3875.5	68.3	1086	21	AAW3819
23	3875.5	68.3	1086	21	AAW3819
24	3875.5	68.3	1086	21	AAW3819
25	3843.5	67.7	1094	21	AAW3819
26	3843.5	67.7	1094	21	AAW3819
27	3843.5	67.7	1094	21	AAW3819
28	3811	67.1	1026	23	AAW3819
29	3804.5	67.0	1039	21	AAW3819
30	3801	67.0	1065	23	AAW3819
31	3748.5	66.0	1039	19	AAW3819
32	3741	65.9	1043	23	AAW3819
33	3670	64.6	1069	23	AAW3819
34	3647.5	64.3	1084	23	AAW3819
35	3628	63.9	1084	23	AAW3819
36	3596	63.3	1080	22	AAW3819
37	3596	63.3	1084	19	AAW3819
38	3590	63.2	1088	23	AAW3819
39	3522	62.0	821	23	AAW3819
40	3476.5	61.2	974	19	AAW3819
41	3473.5	61.2	974	19	AAW3819
42	3348	59.0	955	21	AAW3819
43	3317.5	58.4	958	23	AAW3819
44	3216	56.6	793	21	AAW3819
45	3143	55.4	946	22	AAW3819

ALIGNMENTS

RESULT 1
AAW3819
ID AAW3819 standard; Protein; 1065 AA.
XX AAW3819;
XX 06-JUL-1998 (first entry)
XX Arabidopsis cellulose biosynthetic protein Ath-B.
XX Cellulose; cellulose synthase; RSW1 protein; beta-1,4-glucan;
XX transgenic plant.
XX Arabidopsis thaliana var. Columbia.
XX W09800549-A1.
XX 08-JAN-1998.
XX 24-JUN-1997; 97WO-AU00402.
XX 27-JUN-1996; 96AU-0000699.
XX (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX (AUSU) UNIV AUSTRALIAN NAT.
XX ARIOLI A, Betzner AS, Peng L, Williamson RE;
XX WPI. 1998-086974/08.
XX N-PSDB: AAV06567.
XX DNA encoding cellulose biosynthetic enzyme - useful for manipulation
XX of cellulose and beta-1,4-glucan

XX Claim 29, page 144-149; 207pp; English.
PS
XX
CC This polypeptide is encoded by cDNA clone Ath-B (see AAV06567)
CC that was isolated from an Arabidopsis thaliana cDNA library using
CC PCR primers (see AAT99632-34) based on cellulose synthase RSM1
CC genomic clone 23H12 (see AAV06563) and EST clone AAT20782 (see
CC AAV06562). It is closely related to Arabidopsis cellulose synthase RSM1
CC protein (see AAM33816-17). Claimed nucleic acid molecules (see
CC AAV0562-69) coding for claimed polypeptides (see AAM33816-20 and
CC AAM46202) involved in cellulose biosynthesis can be used to
CC manipulate the cellulose and/or beta-glucan content of transgenic
CC plants. Expression of nucleic acids in the sense orientation
CC increases the level of cellulose and reduces the level of
CC non-crystalline beta-1,4-glucan and starch, providing plants with
CC modified strength and/or shape and/or fibre properties, or having
CC increased resistance to stresses or pests. Antisense, ribozyme or
CC co-suppression molecules can be used to reduce the cellulose
CC content of a transgenic plant, e.g. to improve digestibility or to
CC alter carbon partitioning such that increased carbon is available
CC for growth, rather than deposited as cellulose.
CC
XX
SO Sequence 1065 AA:

Query Match 99.6%; Score 5655; DB 19; Length 1065;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1062; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MESEGTAGKPKMNIYVOTCOICSDNVGKTVDDREFVACDICSFPYCRPCYEYERKDNQ 60
DB 1 MESEGTAGKPKMNIYVOTCOICSDNVGKTVDDREFVACDICSFPYCRPCYEYERKDNQ 60
QY 61 SCPOCKTRYKRLKGPALPGDDEGLADGVEFENYPOKEKISERMGLHTRGKEEM 120
DB 61 SCPOCKTRYKRLKGPALPGDDEGLADGVEFENYPOKEKISERMGLHTRGKEEM 120
QY 121 GEPOYKEVSHNHLPLRTSRODTSGFSAASPERLSYSTIAGKRLPSSDVNOSPNNR 180
DB 121 GEPOYKEVSHNHLPLRTSRODTSGFSAASPERLSYSTIAGKRLPSSDVNOSPNNR 180
QY 181 IYDPVGLGNVAMKERVDGKMKOEKNTGVPSTOASERGVDDIDASTDILADALLNDEA 240
DB 181 IYDPVGLGNVAMKERVDGKMKOEKNTGVPSTOASERGVDDIDASTDILADALLNDEA 240
QY 181 IYDPVGLGNVAMKERVDGKMKOEKNTGVPSTOASERGVDDIDASTDILADALLNDEA 240
DB 181 IYDPVGLGNVAMKERVDGKMKOEKNTGVPSTOASERGVDDIDASTDILADALLNDEA 240
QY 241 ROPLSRKVASIPSSSRINPYRMVIMLRVLICFLHYRITNPVNAFALMTVSVCEIMFAL 300
DB 241 ROPLSRKVASIPSSSRINPYRMVIMLRVLICFLHYRITNPVNAFALMTVSVCEIMFAL 300
QY 301 SWILDOFPKWFYVNRRTYIDRLALRYDRGEGPSQLAVIDFVSTVPLKEPPLVANTYL 360
DB 301 SWILDOFPKWFYVNRRTYIDRLALRYDRGEGPSQLAVIDFVSTVPLKEPPLVANTYL 360
QY 361 STLAADVPAVDKVCYFDDGAAMLSPESLAETSEFAKRWVPCKKXSTIEPRAEWFAAK 420
DB 361 STLAADVPAVDKVCYFDDGAAMLSPESLAETSEFAKRWVPCKKXSTIEPRAEWFAAK 420
QY 421 IDYLDKQVTSFVKDRRAKREYEERKIRINLVSKALCPBEGWVMDGTWPWGNTGD 480
DB 421 IDYLDKQVTSFVKDRRAKREYEERKIRINLVSKALCPBEGWVMDGTWPWGNTGD 480
QY 481 HEGMTQVFLGONGGDAEGNELPRLYVYVREKRPQOHKKKAGAMALYVSAVLTNGEP 540
DB 481 HEGMTQVFLGONGGDAEGNELPRLYVYVREKRPQOHKKKAGAMALYVSAVLTNGEP 540
QY 541 IINLDCDHYINNSKALREAMEFLMDPNLKOVCYVOPORFQIDKNDYANNTYFEDI 600
DB 541 IINLDCDHYINNSKALREAMEFLMDPNLKOVCYVOPORFQIDKNDYANNTYFEDI 600
QY 601 NLRGIDGLOGPYVGTGVFNRTALGYEPPIKVKHKRPSLSKLOGSGRKKNSKAKKS 660
DB 601 NLRGIDGLOGPYVGTGVFNRTALGYEPPIKVKHKRPSLSKLOGSGRKKNSKAKKS 660
QY 661 DKKSGRHTDSTVPFNLDDIEBGVAGCFDDEKALLMSOMLEKRFQSAVFASTLME 720

DB 661 DKKSGRHTDSTVPFNLDDIEBGVAGCFDDEKALLMSOMLEKRFQSAVFASTLME 720
QY 721 NGVPSAPPENLKEAIIHIVISGYEDKSDMWMEIGMIVGSVEDILITGKMHARGRST 780
DB 721 NGVPSAPPENLKEAIIHIVISGYEDKSDMWMEIGMIVGSVEDILITGKMHARGRST 780
QY 781 YCMRKLPAFGKSAFINLSDBNLQVLRWALGSVILFSRHCPYGYNGRLKLEPRAVYN 840
DB 781 YCMRKLPAFGKSAFINLSDBNLQVLRWALGSVILFSRHCPYGYNGRLKLEPRAVYN 840
QY 841 TTYPTITSIPLMWCTLLAVCLTNOPIIPQISNIASIMFLSFLSIFATGILEMRSGV 900
DB 841 TTYPTITSIPLMWCTLLAVCLTNOPIIPQISNIASIMFLSFLSIFATGILEMRSGV 900
QY 901 GIDEMRNEDQFWYIGVSAHLFAVFOGILKVLGIDTNPVTSKASDEDDFAELYFKW 960
DB 901 GIDEMRNEDQFWYIGVSAHLFAVFOGILKVLGIDTNPVTSKASDEDDFAELYFKW 960
QY 961 TLLIPPTLLIYNLVGVVAGVYAINSGVSGPLFGKLFPAFWYIVHLYPELKGIMGR 1020
DB 961 TLLIPPTLLIYNLVGVVAGVYAINSGVSGPLFGKLFPAFWYIVHLYPELKGIMGR 1020
QY 1021 QNRTPTIVVWSVLASIFSLSLWVRIDPFTSRVTGDIIECGINC 1065
DB 1021 QNRTPTIVVWSVLASIFSLSLWVRIDPFTSRVTGDIIECGINC 1065

RESULT 2
AAC48733
ID AAC48733 standard; Protein; 1065 AA.
XX
AC AAG48733;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61573.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
PN BP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 9905-0121825.
PR 05-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 28-APR-1999; 9905-0130891.
PR 30-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 07-MAY-1999; 9905-0132487.
PR 11-MAY-1999; 9905-0132863.
PR 11-MAY-1999; 9905-0134256.

PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134378.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 17-JUN-1999; 9905-0139453.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 23-JUL-1999; 9905-0145087.
PR 23-JUL-1999; 9905-0145089.
PR 23-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.

PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147028.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147312.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148365.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0156599.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.

```

PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161921.
PR 28-OCT-1999; 9905-0161922.
PR 28-OCT-1999; 9905-0161923.
PR 29-OCT-1999; 9905-0162142.

```

```

Query Match 99.6%; Score 5654; DB 21; Length 1065;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1062; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 MEESEGTACKPMKNIVPOTQICSDNVGKTVGDREPVACDICSFPVCRPCYEYERKDGNO 60
DB 1 MEESEGTACKPMKNIVPOTQICSDNVGKTVGDREPVACDICSFPVCRPCYEYERKDGNO 60
QY 61 SCPOCKTRYKRLKGSFAIPGDKDEDEGLADBGVTEFVNPPOKEKISEMNLGWHLTRGGEEM 120
DB 61 SCPOCKTRYKRLKGSFAIPGDKDEDEGLADBGVTEFVNPPOKEKISEMNLGWHLTRGGEEM 120
QY 121 GEPOYDKEVSHNHLPRLTSGRODTSGEFSASPERLSVSTIAGKRLPYSSDVNOGPNRR 180
DB 121 GEPOYDKEVSHNHLPRLTSGRODTSGEFSASPERLSVSTIAGKRLPYSSDVNOGPNRR 180
QY 181 IYDPVGLGNVAMKERDGMKMKOEKNTGPVSTQASERGVDDIDASTDILADALLNDEA 240
DB 181 IYDPVGLGNVAMKERDGMKMKOEKNTGPVSTQASERGVDDIDASTDILADALLNDEA 240
QY 241 ROPLSRKVSIPSSSRINPYRMVIMLRVLIICLFHYRITNPVPAFALMLVSVCEIEMFAL 300
DB 241 ROPLSRKVSIPSSSRINPYRMVIMLRVLIICLFHYRITNPVPAFALMLVSVCEIEMFAL 300
QY 301 SWLIDOPKWFVPVNRETYLDRLALRYDREGSPQLAAVDIEVSTVDPLEKPLVTANTYL 360
DB 301 SWLIDOPKWFVPVNRETYLDRLALRYDREGSPQLAAVDIEVSTVDPLEKPLVTANTYL 360
QY 361 SIILAVYYPVDKVCYVFDDGAAMLSESLAETSEFARKVWPFCKKTSIEPRABEWYFAAK 420
DB 361 SIILAVYYPVDKVCYVFDDGAAMLSESLAETSEFARKVWPFCKKTSIEPRABEWYFAAK 420
QY 421 IDYLKDKVOVSFYKDDRRAMKREYEERKIRINALVSKALCPREGWVWODTPPGNNTG 480
DB 421 IDYLKDKVOVSFYKDDRRAMKREYEERKIRINALVSKALCPREGWVWODTPPGNNTG 480
QY 481 HPGMIOVFLGONGGLDAEGNELPRLVYVSREKRPFQHHKKAGAMNALVVSALVINGPF 540
DB 481 HPGMIOVFLGONGGLDAEGNELPRLVYVSREKRPFQHHKKAGAMNALVVSALVINGPF 540
QY 541 IINLDDHYIINNSKALREAMCFIAMDNLKQVCYVQFPRQFQDIDKNDNRANNNTYFPI 600
DB 541 IINLDDHYIINNSKALREAMCFIAMDNLKQVCYVQFPRQFQDIDKNDNRANNNTYFPI 600
QY 601 NLNGLDIGIOPVYVGTGVFNRTALGYEPPIKVKKKPSLSKLGGSKKKSKAKKES 660
DB 601 NLNGLDIGIOPVYVGTGVFNRTALGYEPPIKVKKKPSLSKLGGSKKKSKAKKES 660
QY 661 DKKSGRHTDSTVPVFNLDIEGVEGAGFDEKALLMSOMLEKRRFGSAVFASTLME 720
DB 661 DKKSGRHTDSTVPVFNLDIEGVEGAGFDEKALLMSOMLEKRRFGSAVFASTLME 720
QY 721 NGGVPSATPENLTKAIIHIVISGVEYEDKSPMGMEIOWIYGSYVEDIILTGKMAARWRST 780
DB 721 NGGVPSATPENLTKAIIHIVISGVEYEDKSPMGMEIOWIYGSYVEDIILTGKMAARWRST 780
QY 781 YCMPIKPAKGSAPINLSDRLNOVLKMAALGSVEILSRHCPIMYGNGRLKFLERAAVYA 840
DB 781 YCMPIKPAKGSAPINLSDRLNOVLKMAALGSVEILSRHCPIMYGNGRLKFLERAAVYA 840
QY 841 TTIYPTISIPLLMYCTLLAVCLFTNOPIIPIQISINIASIWFLSLFISFATGILEMRSGV 900

```

```

DB 841 TTIYPTISIPLLMYCTLLAVCLFTNOPIIPIQISINIASIWFLSLFISFATGILEMRSGV 900
QY 901 GIDEMWRNPOFWVIGVSAHLPAVFGCIIKVLAGIDTNTVTSKASDEGDAELYLFRW 960
DB 901 GIDEMWRNPOFWVIGVSAHLPAVFGCIIKVLAGIDTNTVTSKASDEGDAELYLFRW 960
QY 961 TTIILIPPTLLIYNLGVAVGVSAYAINSGVSGPLGKLFPAFWIVHILYPLKGLMR 1020
DB 961 TTIILIPPTLLIYNLGVAVGVSAYAINSGVSGPLGKLFPAFWIVHILYPLKGLMR 1020
QY 1021 QNRTPTIVVWSVLASIFSLSLWVRIDPFTSRVTGPDILIECGINC 1065
DB 1021 QNRTPTIVVWSVLASIFSLSLWVRIDPFTSRVTGPDILIECGINC 1065

```

```

RESULT 3
ABB93356
ID ABB93356 standard; Protein; 1065 AA.
XX
AC ABB93356;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 2567.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB ) BAYER AG.
XX
PI Tietjen K, Weidner M.
XX
DR WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comparing aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
PS Claim 5; SEQ ID NO 2567; 261bp + Sequence Listing; English.
XX
XX
CC The invention relates to identifying target proteins
CC (ABB90730-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
SQ Sequence 1065 AA;

```

```

Query Match 99.6%; Score 5654; DB 23; Length 1065;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1062; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 MEESEGTACKPMKNIVPOTQICSDNVGKTVGDREPVACDICSFPVCRPCYEYERKDGNO 60
DB 1 MEESEGTACKPMKNIVPOTQICSDNVGKTVGDREPVACDICSFPVCRPCYEYERKDGNO 60
QY 61 SCPOCKTRYKRLKGSFAIPGDKDEDEGLADBGVTEFVNPPOKEKISEMNLGWHLTRGGEEM 120

```

Db 61 SCPOCKTRYKRLKGGSPAI PGDKEDGLADEGTVEFNYPQKEKISERMLGWHLTRKGGEEM 120
QY 121 GEPQDKEVSHNHLRLTSRODTSGEFSAASPERLSVSSTTNGKRLPSPVNO SPNRR 180
Db 121 GEPQDKEVSHNHLRLTSRODTSGEFSAASPERLSVSSTTNGKRLPSPVNO SPNRR 180
QY 181 IVDPVGLGNVAMKEKREVDGWMKMKQEKNTGTPVSTQASERGVVIDIDASTDILADEALLNDEA 240
Db 181 IVDPVGLGNVAMKEKREVDGWMKMKQEKNTGTPVSTQASERGVVIDIDASTDILADEALLNDEA 240
QY 241 RQPLSRKVSIPSSRRINPYRMVIMLVILCLFLHRTINPVNAFALMVSYICEIWFAL 300
Db 241 RQPLSRKVSIPSSRRINPYRMVIMLVILCLFLHRTINPVNAFALMVSYICEIWFAL 300
QY 301 SWILDQFPKMFVYNETYLDRALRYDRGEPSSOLAADIENSTVDPLKEPPLVYANVYL 360
Db 301 SWILDQFPKMFVYNETYLDRALRYDRGEPSSOLAADIENSTVDPLKEPPLVYANVYL 360
QY 361 SILAVDYVDKVCYVFPDGAAMLSEFSLAETSEFARKWVPECKKYSIEPRAPEWYFAK 420
Db 361 SILAVDYVDKVCYVFPDGAAMLSEFSLAETSEFARKWVPECKKYSIEPRAPEWYFAK 420
QY 421 IDYLKDKVQTSFVNDKRRAMKREYEEFKIRINALVSKALCPREGEVVMODGTPWPGNNTGD 480
Db 421 IDYLKDKVQTSFVNDKRRAMKREYEEFKIRINALVSKALCPREGEVVMODGTPWPGNNTRD 480
QY 481 HPGMIQVFLGONGGLDAGNELPRLVYVSRKRPFGOHKKGANNALVRSVAVLTNGPF 540
Db 481 HPGMIQVFLGONGGLDAGNELPRLVYVSRKRPFGOHKKGANNALVRSVAVLTNGPF 540
QY 541 ILNLDCDHY INNSKALREAMCFLMDPNLKGQVCYVQFQRFQDIDKNRYANRNTVFEDI 600
Db 541 ILNLDCDHY INNSKALREAMCFLMDPNLKGQVCYVQFQRFQDIDKNRYANRNTVFEDI 600
QY 601 NLRGIDGIGQPVYVGTGCVFNRTALXGEPRLKVNKKRPSLSKLCGSSRKNSAKKES 660
Db 601 NLRGIDGIGQPVYVGTGCVFNRTALXGEPRLKVNKKRPSLSKLCGSSRKNSAKKES 660
QY 661 DKKSGRHTDSTVPFNLDDIEEGEGAGFDEKALLSOMSELEKRFQGSAYFVASTLME 720
Db 661 DKKSGRHTDSTVPFNLDDIEEGEGAGFDEKALLSOMSELEKRFQGSAYFVASTLME 720
QY 721 NGVPPSATPENLKEALIHVISCYGEDKSDMCMETIGWYGVSTEDILTGFKNHARGWSI 780
Db 721 NGVPPSATPENLKEALIHVISCYGEDKSDMCMETIGWYGVSTEDILTGFKNHARGWSI 780
QY 781 YCMPLRAFKGSAPINLSDRNLQVLRMALGSEIIFSRNCRPTWYGNRKLKLERFAYVN 840
Db 781 YCMPLRAFKGSAPINLSDRNLQVLRMALGSEIIFSRNCRPTWYGNRKLKLERFAYVN 840
QY 841 TTIYPTISIPILMYCTLLAVCLFTNQFIIPQISNIASTWFLSLFSTIRATGILEMRWGV 900
Db 841 TTIYPTISIPILMYCTLLAVCLFTNQFIIPQISNIASTWFLSLFSTIRATGILEMRWGV 900
QY 901 GIDEMWRNEQFWITGVSANHLFAVQGIKVLGIDTFTYVTSKASDEGDGFALYLTKM 960
Db 901 GIDEMWRNEQFWITGVSANHLFAVQGIKVLGIDTFTYVTSKASDEGDGFALYLTKM 960
QY 961 TLLILPPTLLILVNLGVVAGVSTAINSGYSQWGLFEKLPFAFVYIHLVPLKGLMGR 1020
Db 961 TLLILPPTLLILVNLGVVAGVSTAINSGYSQWGLFEKLPFAFVYIHLVPLKGLMGR 1020
QY 1021 ONRPTIYVWMSVLLASIFSLMWRIDPFTSRVTPDILEGGINC 1065
Db 1021 ONRPTIYVWMSVLLASIFSLMWRIDPFTSRVTPDILEGGINC 1065

RESULT 4
ID AAG48734
AC AAG48734;
XX AAG48734;
XX

DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61574.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137528.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.

PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139617.
 PR 22-JUN-1999; 99US-0139699.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142290.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144684.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 06-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 09-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 10-AUG-1999; 99US-0147935.
 PR 11-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152263.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 98.6%; Score 5597; DB 21; Length 1054;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1051; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 12 MNIVPQTQICSDNKGTVDDGFVACDICEFVPCRPCEYERKDNQSCPOCKTRYR 71
 DB 1 MNIVPQTQICSDNKGTVDDGFVACDICEFVPCRPCEYERKDNQSCPOCKTRYR 60

QY 72 LKSPAIIPDCKDEGLADGTYEFNYPQEKISERLGNHLTRGKEENGEPQYDKVEVH 131
 DB 61 LKSPAIIPDCKDEGLADGTYEFNYPQEKISERLGNHLTRGKEENGEPQYDKVEVH 120

QY 132 NMLPRLTSRODTSGEESAASPERLSVSTIAGCKRLPYSSDVNOSPNNRRTVPVGLGNYA 191
 DB 121 NMLPRLTSRODTSGEESAASPERLSVSTIAGCKRLPYSSDVNOSPNNRRTVPVGLGNYA 180

QY 192 MKERVDGWMKQEKNTGPVSTQASERGGVDIDASTDILADEALLNDEAROPLSRKVSIP 251
 DB 181 MKERVDGWMKQEKNTGPVSTQASERGGVDIDASTDILADEALLNDEAROPLSRKVSIP 240

QY 252 SSRINFRVYIMLRVILCLFLHYRTTNVYPAFALMIVSVICETIFALSWILDOPPKVF 311
 DB 252 SSRINFRVYIMLRVILCLFLHYRTTNVYPAFALMIVSVICETIFALSWILDOPPKVF 311

Db 241 SSRINPFRMVLRLVILCLFLHYRITNPVPAFLMLVSYICEIWFALSWILDQFPKMF 300
OY 312 PUNRETYDLRLALRYDREGPSOLAANDIEVSTDPDKKEPPLVANTVLTSLADYPRDK 371
Db 301 PUNRETYDLRLALRYDREGPSOLAANDIEVSTDPDKKEPPLVANTVLTSLADYPRDK 360
OY 372 VSCYVFDGAAMLSFESIAETSEFARKVVPFCCKYSIEPRAPENYFAAKIDYLDKQVTS 431
Db 361 VSCYVSDGAAMLSFESIAETSEFARKVVPFCCKYSIEPRAPENYFAAKIDYLDKQVTS 420
OY 432 FVKDRAMKREYEEFKIRINALVSKALCKPEEGVMODGTWPNGNTGDHGMIOVFLGQ 491
Db 421 FVKDRAMKREYEEFKIRINALVSKALCKPEEGVMODGTWPNGNTGDHGMIOVFLGQ 480
OY 492 NGGDAEGNELPRVLYVYSREKRPFGOHKKAGANALVRSVAVLTNGFTINLDCDHTIN 551
Db 481 NGGDAEGNELPRVLYVYSREKRPFGOHKKAGANALVRSVAVLTNGFTINLDCDHTIN 540
OY 552 NSKALREAMCFIAMPNLSKQCYVOFPORFGIDKNDRYANRNTVFPIINLRGLDGIOP 611
Db 541 NSKALREAMCFIAMPNLSKQCYVOFPORFGIDKNDRYANRNTVFPIINLRGLDGIOP 600
OY 612 VYVGTGCVFNRTALYGEPPIKVKKRPSLLSKLGGSRKKNSAKKESDKKSGRHTDS 671
Db 601 VYVGTGCVFNRTALYGEPPIKVKKRPSLLSKLGGSRKKNSAKKESDKKSGRHTDS 660
OY 672 TVPVENLDDIEGVEGAFDEDEKALMSQMLEKRFQGSANFVASTLMENGVPSAPPE 731
Db 661 TVPVENLDDIEGVEGAFDEDEKALMSQMLEKRFQGSANFVASTLMENGVPSAPPE 720
OY 732 NLLEKAIHVISCYGEDSKDMGMEIGWYGTEDILTFGKMHARGRSIYCMPLPAPKG 791
Db 721 NLLEKAIHVISCYGEDSKDMGMEIGWYGTEDILTFGKMHARGRSIYCMPLPAPKG 780
OY 792 SAPINLSDRNLQVLRMALGSVEILFSRHCPIMYGNRKLKLERPAAVNTTPIYTSIPL 851
Db 781 SAPINLSDRNLQVLRMALGSVEILFSRHCPIMYGNRKLKLERPAAVNTTPIYTSIPL 840
OY 852 LMCTCLAVCLFTNOFTIIPQISNASTIMELSLFISTFATGILEMWSGVGIDEMWRNQF 911
Db 841 LMCTCLAVCLFTNOFTIIPQISNASTIMELSLFISTFATGILEMWSGVGIDEMWRNQF 900
OY 912 WVIQGVSAHLFAVFOGILKVLAGIDNTFTVTSKASDEGDAELIYLFKWTLLLPPTLL 971
Db 901 WVIQGVSAHLFAVFOGILKVLAGIDNTFTVTSKASDEGDAELIYLFKWTLLLPPTLL 960
OY 972 IVNLVGVAVGVSYAINSQYQSMGPLFGKLFAPFVVIYHLVLPFLGLMGRONRTPYVVM 1031
Db 961 IVNLVGVAVGVSYAINSQYQSMGPLFGKLFAPFVVIYHLVLPFLGLMGRONRTPYVVM 1020
OY 1032 SVLASTIFSLMVRIDPFTSRVTPGPDILEGGINC 1065
Db 1021 SVLASTIFSLMVRIDPFTSRVTPGPDILEGGINC 1054

RESULT 5
AAG48735
ID AAG48735 standard; Protein: 959 AA.
XX AAG48735;
AC
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61575.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
OS
XX
PN EPI033405-A2.
XX

PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138099.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0143297.
PR 13-JUL-1999; 9905-0143477.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145313.
PR 27-JUL-1999; 9905-0145318.
PR 27-JUL-1999; 9905-0145319.
PR 28-JUL-1999; 9905-0145351.
PR 02-AUG-1999; 9905-0145386.
PR 02-AUG-1999; 9905-0146388.
PR 03-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 06-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 07-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 22-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.

PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161982.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

Query Match 89.2%; Score 5066; DB 21; Length 959;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 956; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 107 MCGMHLTRKGEEMGEPQYDKESVHNHLRLTSLRDTSGEFSNAPSRLTSSVSTIAGRR 166
DB 1 MCGMHLTRKGEEMGEPQYDKESVHNHLRLTSLRDTSGEFSNAPSRLTSSVSTIAGRR 60
QY 167 LPTSSDVNOSPNNRITDVPGLGVNWKKEVDGMMKOEKNTGPVSTQASERGGVDIDS 226
DB 61 LPTSSDVNOSPNNRITDVPGLGVNWKKEVDGMMKOEKNTGPVSTQASERGGVDIDS 120
QY 227 TDLTADLNDENARQPLSRKYSIPSSRIINPYRMVIMLRVILICLFLAHYRITNPVNAFA 286
DB 121 TDLTADLNDENARQPLSRKYSIPSSRIINPYRMVIMLRVILICLFLAHYRITNPVNAFA 180
QY 287 LMLVSVICETWFLSNIIDQPKWFPVNBETYLRLALRYDRGEPSOLAADVSTVD 346
DB 181 LMLVSVICETWFLSNIIDQPKWFPVNBETYLRLALRYDRGEPSOLAADVSTVD 240
QY 347 PLKEPPLVANTVLTSLAVDYVPDKSCYVPFDGAMLSFEESLAESEFARKWVPCKKY 406
DB 241 PLKEPPLVANTVLTSLAVDYVPDKSCYVPFDGAMLSFEESLAESEFARKWVPCKKY 300
QY 407 SIEPRAPENYFAKIDYLDKQVTSFVKDRRAKREYEFKIRINALVSKALCPEEGVY 466
DB 301 SIEPRAPENYFAKIDYLDKQVTSFVKDRRAKREYEFKIRINALVSKALCPEEGVY 360
QY 467 MDGTPWPGNNTGDHDMGTOVFLGONGGLDAGCNELPRLVYYSREKRPFGOHKKRAGAN 526
DB 361 MDGTPWPGNNTGDHDMGTOVFLGONGGLDAGCNELPRLVYYSREKRPFGOHKKRAGAN 420
QY 527 ALYRVSALVTNGFFIILNDCDHYINNSKALBEMCMCLAMPNIGKQYCYQFQRPBGIDK 586
DB 527 ALYRVSALVTNGFFIILNDCDHYINNSKALBEMCMCLAMPNIGKQYCYQFQRPBGIDK 586

Db 421 ALVRSAVLNTPGPTLNLDCDHYINNSKALREAMCFMLDMPNLGKQVCYVQPPQREIDK 480
 QY 587 NDRKANRTVFEDINLRGLDGIQGVYVGTGVNFRALTYGEPPIKXHKHKKPSLSLKC 646
 Db 481 NDRKANRTVFEDINLRGLDGIQGVYVGTGVNFRALTYGEPPIKXHKHKKPSLSLKC 540
 QY 647 GGSRRKNSKAKKESDKKSGRHTDSTVFPNLDDIEEGEGAGFDEKALLMSQMSLEKR 706
 Db 541 GGSRRKNSKAKKESDKKSGRHTDSTVFPNLDDIEEGEGAGFDEKALLMSQMSLEKR 600
 QY 707 FGQSAVFASVLMENGVPPSATPENLLKEAIVHISCYEDKSDMGMEIGWYGVTEDI 766
 Db 601 FGQSAVFASVLMENGVPPSATPENLLKEAIVHISCYEDKSDMGMEIGWYGVTEDI 660
 QY 767 LTGFEMHARGRSITCMKPLPAFKGSAPINLSDRLNOVLKRNALGVELFSGHCPIMGY 826
 Db 661 LTGFEMHARGRSITCMKPLPAFKGSAPINLSDRLNOVLKRNALGVELFSGHCPIMGY 720
 QY 827 NGRLEKLEBFAVYVNTTIPITSIPILMYCTLLAVCLFTNOFTIPOISNASTWFLSFLS 886
 Db 721 NGRLEKLEBFAVYVNTTIPITSIPILMYCTLLAVCLFTNOFTIPOISNASTWFLSFLS 780
 QY 887 IFATGILEMRWSGVGIDEMWRNEOPWVIGVSAHLFAVFOGILKVLACIDINFTVTSKAS 946
 Db 781 IFATGILEMRWSGVGIDEMWRNEOPWVIGVSAHLFAVFOGILKVLACIDINFTVTSKAS 840
 QY 947 DEDGDFALYLFKMTLLIPTTLIVNLGVVAGVSAINSYGOSMGPLGKLFEPFV 1006
 Db 841 DEDGDFALYLFKMTLLIPTTLIVNLGVVAGVSAINSYGOSMGPLGKLFEPFV 900
 QY 1007 IVHLVPLKGLMGRNRTPTIVVWSVLASIFSLMWRIDPFTSRVGPDLLEGGINC 1065
 Db 901 IVHLVPLKGLMGRNRTPTIVVWSVLASIFSLMWRIDPFTSRVGPDLLEGGINC 959

RESULT 6
 AAY84108
 ID AAY84108 standard; Protein: 1077 AA.

AC AAY84108;
 DT 03-JUL-2000 (first entry)
 DE Amino acid sequence of a maize cellulose synthase.
 DE Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
 KM transgenic plant; plant breeding marker.
 OS Zea mays.
 FH Key Location/Qualifiers
 FT Misc-difference 494 /note= "encoded by MGC"
 PN WO200009706-A2.
 PD 24-FEB-2000.
 PE 16-AUG-1999; 99WO-US18760.
 PR 17-AUG-1998; 98US-0096822.
 PA (PION-) PIONEER HI-BRED INT INC.
 PI Dhugga KS, Helentjaris TG, Bowen BA, Wang X;
 DR WPI; 2000-224343/19.
 DR N-PSDB; AA299494.
 PT New genes which encode maize cellulose synthase polypeptides in plants
 PT useful for modulating the expression of cellulose synthase in plants
 PT and to produce transgenic plants expressing the novel protein
 XX

Claim 15; Page 95-97; 119pp; English.

The present sequence represents a maize cellulose synthase polypeptide.
 The cellulose synthase can be used for the improvement of stalk quality
 for improved stand or silage. It also provides an increased concentration
 of cellulose in the pericarp, hardening the kernel and improving its
 handling ability. The sequences are used to produce transgenic plants
 and seeds expressing the cellulose synthase. The polynucleotide is
 used for modulating, preferably increasing, the level of the synthase
 in a plant cell. The plants are preferably monocots. The polynucleotide
 is also used as a probe or primer in the detection quantitation or
 isolation of gene transcripts. The probes are useful in detecting
 deficiencies in the level of mRNA in screenings for desired transgenic
 plant, for detecting mutations in the gene, for monitoring upregulation
 of expression or changes in enzyme activity in screening assays of
 compounds, for detection of any number of allelic variants of the gene,
 or for use as molecular markers in plant breeding programs. The
 isolated nucleic acids of the present invention can also be used for
 recombinant expression of their encoded polypeptides or for use as
 immunogens in the preparation and/or screening of antibodies. The
 proteins can be employed in assays for enzyme agonists or antagonists
 of enzyme function or for use of immunogens or antigens to obtain
 antibodies specifically immunoreactive with a protein.

SQ Sequence 1077 AA;
 Query Match 80.5%; Score 4572; DB 21; Length 1077;
 Best Local Similarity 79.4%; Pred. No. 0;
 Matches 861; Conservative 81; Mismatches 112; Indels 30; Gaps 10;

QY 4 EGERTAG-KPMKNIVYQTCICSDNNGYVDREFAICISFPVPCYERKQDNGNSC 62
 Db 2 EGDADGVASGRGGGQVQICGQVGTGTAEGDVFACVCGFPYCPCTEYERKDGTOAC 61
 QY 63 PCKKTRKRLKSPALPGDKDEGLADEGVEFNT-----PQKEKISERMLGWHLTRGK 116
 Db 62 PCKKTRKRLKSPALPGDKDEGLADEGVEFNT-----PQKEKISERMLGWHLTRGK 118
 QY 117 GEEMGEPOYDK-----EVSHNLPRLTSQDPTSGEFSASPERLSVSTIAGCK 165
 Db 119 SGDVGPRKYDSGEIGLTRYDSGEIPRGYIPVSTNSQ-ISGEIPGASPHHMSPGNIGK 177
 QY 166 RLPSYSDVNOSENRIYDPVGVGNVAMKERVDMGKMKQKNTGP---STOASPRGCV 221
 Db 178 RAPFPY-VNHSPPNRSERSSGIGNVAMKERVDMGKMKQKNTGP---STOASPRGCV 236
 QY 222 DIDASTDILADEALLNDEARQPLSRKYSIPSSRINPYRMVIMLRVLVLCFLHYRITNPV 281
 Db 237 DIDASTDYNMEDALLNDETROPPLSRKYPVLPSSRINPYRMVIMLRVLVLCFLHYRITNPV 296
 QY 282 PNAFALMLVSYICEIWFALSWILDQFPKMFVFNRETYYIDRLALRDRGGEPSQLAAVDIF 341
 Db 297 RNAYPLMLLSYICEIWFALSWILDQFPKMFVFNRETYYIDRLALRDRGGEPSQLAAVDIF 356
 QY 342 VSTVDPLKEPPLVNTANTLSTILAVDYPVDKVSICYFDDGAAMLSPESLAETSEFARKVVP 401
 Db 357 VSTVDPMKEPPLVNTANTLSTILAVDYPVDKVSICYFDDGAAMLSPESLAETSEFARKVVP 416
 QY 402 FCKRYSIEPRAPEMYFAAKIDYLDKDYQTSFVKDRAMKREYEERKIRINALVSKALCP 461
 Db 417 FVKRYNIEPRAPEMYFQKIDYLDKDYQTSFVKDRAMKREYEERKIRINALVSKALCP 476
 QY 462 EEGVMQDGTWPGNNTGDHFGMTQVFLGQNGCGLDABENELPRLVYVSREKRPQGHKK 521
 Db 477 EEGVMQDGTWPGNNTGDHFGMTQVFLGQNGCGLDABENELPRLVYVSREKRPQGHKK 536
 QY 522 AGAMNALVRSVAVLNGPFIINLDCDHYINNSKALREAMCFMLDMPNLGKQVCYVQPPQRE 581
 Db 537 AGAMNALVRSVAVLNGPFIINLDCDHYINNSKALREAMCFMLDMPNLGKQVCYVQPPQRE 596
 QY 582 DGIDNDRKANRTVFEDINLRGLDGIQGVYVGTGVNFRALTYGEPPIKXHKHKKPSL 641
 Db 597 DGIDNDRKANRTVFEDINLRGLDGIQGVYVGTGVNFRALTYGEPPIKXHKHKKPSL 654

QY 642 LSLKCGSRKKSKAKKESDKKSGRHTDSTVPEVFNLDIEEGVAGFDEDEKALLMSQ 701
DB 655 LSSLGCG-RKKASKSKGSDKKSKQHVDSSVFNLEDIEEGVAGFDEDEKALLMSQ 713
QY 702 SLEKRRGQSAFVASTLMENGVPSPATPENLKEAIVHISGVEDKSDMGMEIGWYGS 761
DB 714 SLEKRRGQSAFVASTLMENGVPSPATPENLKEAIVHISGVEDKSDMGMEIGWYGS 773
QY 762 VTEDILTFGRMHARGRSIYCMPLKPAFGSAPINISDRNOYLKRNALGSVELLFSSHCP 821
DB 774 VTEDILTFGRMHARGRSIYCMPLKPAFGSAPINISDRNOYLKRNALGSVELLFSSHCP 833
QY 822 IMVYNGRLKFLERFAYVNTTYPITSIPILMYCTLLAVCLTNOFIIPDISNIASIMFL 881
DB 834 LMVYGGRLKFLERFAYVNTTYPILSIPILMYCTLLAVCLTNOFIIPDISNIASIMFL 893
QY 882 SLEFLSFAFGILEMRSGVIGIDEMRNEDQFVWIGISAHLEFAVFOGLIKVLAGIDINFTV 941
DB 894 SLEFLSFAFGILEMRSGVIGIDEMRNEDQFVWIGISAHLEFAVFOGLIKVLAGIDINFTV 953
QY 942 TSKASDEDDGFAELTYFKMTTLLIPTTLLIIVLVGVAGVSAINSGVSGMGLKFL 1001
DB 954 TSKASDEDDGFAELTYFKMTTLLIPTTLLIIVLVGVAGVSAINSGVSGMGLKFL 1013
QY 1002 FAFWIVHLVPLKGLMGRNRTPIVWVSVLASIFSLMWRIDPFTSRVGPDIIEC 1061
DB 1014 FAFWIVHLVPLKGLMGRNRTPIVWVSVLASIFSLMWRIDPFTSRVGPDIIEC 1073
QY 1062 GINC 1065
DB 1074 GINC 1077

RESULT 8

AAU10496
ID AAU10496 standard; Protein; 1079 AA.

XX AAU10496:

DT 14-FEB-2002 (first entry)

DE Corn cellulose synthase Cqrae19/cesa-9.

KM Corn; cellulose synthase; Cdpgs45; cesa-3; Cqrae19; cesa-9;

KW stalk quality; improved stand; silage; pericarp; kernel hardening;

XX handling ability; transgenic plant; immunogen.

OS Zea mays.

XX WO200179516-A2.

XX 25-OCT-2001.

XX 12-APR-2001; 2001WO-US11951.

XX 14-APR-2000; 2000US-0550483.

XX (PION-) PIONEER HI-BRED INT INC.

XX Dhuga KS, HeJentharis TG;

XX WPI: 2002-041338/05.

XX N-PSDB: AAS16458.

XX New cellulose synthase polypeptides and polynucleotides, useful in

XX improving stalk quality or silage, and in increasing concentration of

XX cellulose in the pericarp, hardening the kernel for improved handling

XX ability

XX Claim 3; Page 85-87; 88pp; English.

XX The invention relates to isolated nucleic acids encoding two cellulose

XX synthase proteins from corn, Cdpgs45 (cesa-3) and Cqrae19 (cesa-9).

CC Also disclosed are a recombinant expression cassette comprising the
CC polynucleotide (operably linked to a promoter) a host cell comprising
CC the recombinant expression cassette and a transgenic plant comprising
CC the recombinant expression cassette. The nucleic acid is useful in
CC the improvement of stalk quality for improved stand or silage, and in
CC the improved concentration of cellulose in the pericarp, hardening
CC the kernel, and thus improving its handling ability. The nucleic acids
CC may also be used as probes or amplification primers in the detection,
CC quantification or isolation of gene transcripts, as probes in detecting
CC deficiencies in the level of mRNA, for detecting gene mutations or
CC allelic variants, for monitoring up regulation of expression or changes
CC in enzyme activity in screening assays, for site directed mutagenesis,
CC in sense or antisense suppression of one or more genes in a host
CC cell, tissue or plant. The polypeptides may be used in assays for
CC enzyme agonists or antagonists, and as immunogen or antigen to
CC obtain antibodies specifically immunoreactive with the protein.
CC The present sequence represents a corn cellulose synthase of the
CC invention.

XX Sequence 1079 AA:

Query Match 80.4%; Score 4564; DB 23; Length 1079;

Best Local Similarity 79.1%; Pred. No. 0;

Matches 858; Conservative 86; Mismatches 111; Indels 30; Gaps 11;

4 EGETAG-KPMKNIVPOTCOICSDNKGTVGDGRFVACDIOCPVCRPCYEERKDNOSC 62

2 EGDADGVKSGRRGGVQCIGDGVGTAGDFVACDVGFPVCRPCYEERKDNOSC 61

63 PCKTRYKRLKGSFAIPGDKDEGLADEGVENFP-----OKERISERMLQMHLLRGR 116

62 PCKKNYKRLKGSFAIPGDKDEGLADEGVENFP-----OKERISERMLQMHLLRGR 120

117 GEEMGEPOYDK-----EVSNNHLPRLTSKODTSGEESASPERLSVSTTAGCK 165

121 SGDVGRPKYDSGEIGLTKYDSGEIPRKYIPSVTNSQ-ISGEIIGASPDHMMSPTEINIGR 179

166 RLPLYS-SDVNSQSPNRRLVDPVGLGVNAKREYVGMKMKOKKNGPV---STOASERGV 221

180 RAEPFYNNHSSNPSREPSGV--GVNAKREYVGMKMKOKKNGPV---STOASERGV 237

222 -DIDASTDLADEALLNDEARQPLSRKVSIPSSRINPYRWIMLRVILCLFLHYRTNP 280

238 GDIDASTDVMMEDALLNDEARQPLSRKVSIPSSRINPYRWIMLRVILCLFLHYRTNP 297

281 VPAFAIMLVSVICETIPALSWILDQPKKFPVYNRETYDLRLALRDREGEPSQLAVID 340

298 VRNAVPLMLSVICETIPALSWILDQPKKFPVYNRETYDLRLALRDREGEPSQLAVID 357

341 FVSTVPLKEPPLVTANTVLSILAVDYPVQVSCYVDGAAALSFESLAETSEFARKW 400

358 FVSTVPLKEPPLVTANTVLSILAVDYPVQVSCYVDGAAALSFESLAETSEFARKW 417

401 PFCKKYSIEBRAPEWYFAAKIDYLDKQVQTSFVKDRAMKREYEERFIRINALVSKALC 460

418 PFVKKYNIEBRAPEWYFAAKIDYLDKQVQTSFVKDRAMKREYEERFIRINALVSKALC 477

461 PEEGWYMODGTPMPGNNTGDHPGMIQVFLQNGGLDAEGNELPRLVYVSEKRRPGOHK 520

478 PEEGWYMODGTPMPGNNTGDHPGMIQVFLQNGGLDAEGNELPRLVYVSEKRRPGOHK 537

521 KAGAMNALVSVAVLNGPFLMLDCDHYTNNKALREACFLMDPRLGQVGYVOPOR 580

538 KAGAMNALVSVAVLNGPFLMLDCDHYTNNKALREACFLMDPRLGQVGYVOPOR 597

581 FDGIDRNDRYANNTVEFFDINLGLDIOGPVYVGTGCVFNRTALGYEPPIKVKRKS 640

598 FDGIDRNDRYANNTVEFFDINLGLDIOGPVYVGTGCVFNRTALGYEPPI--KQKKG 655

641 LSLKCGSRKKSKAKKESDKKSGRHTDSTVPEVFNLDIEEGVAGFDEDEKALLMSQ 700

656 LSSLGCG-RKKASKSKGSDKKSKQHVDSSVFNLEDIEEGVAGFDEDEKALLMSQ 714

QY 701 MSLEKRGASAVVASTLMENGVPSPATPENLKEAIVHISCGYEDKSDMGMEIGMIYGS 760
 DB 715 MSLEKRGASAAVASTLMYGVGPSPATPENLKEAIVHISCGYEDKSDMGMEIGMIYGS 774
 QY 761 SVEDILTGFKMARGRSITCMKPLAFKGSAPINISDNLNOLRALGSEVLEIRHC 820
 DB 775 SVEDILTGFKMARGRSITCMKPLAFKGSAPINISDNLNOLRALGSEVLEIRHC 834
 QY 821 PIWYNGRLKFLERFAVNTTITPIYISPLMLCTLAACLEFNPIIPQISNIASIMF 880
 DB 835 PLWYGVGRKFLERFAVNTTITPIYISPLMLCTLAACLEFNPIIPQISNIASIMF 894
 QY 881 LSLPLSFATGILLEMRSVGVDEWNEQFWYIGVSAHLFAVFOGILKVLADITNET 940
 DB 895 ISLPLSFATGILLEMRSVGVDEWNEQFWYIGVSAHLFAVFOGILKVLADITNET 954
 QY 941 VTASDEDEDFALVYFKMTLLIPTTLITNLVGVVSAVINSQSGMPLGAL 1000
 DB 955 VTASDEDEDFALVYFKMTLLIPTTLITNLVGVVSAVINSQSGMPLGAL 1014
 QY 1001 FFAFWIVHLVLPFLKGLMGHONRPTIVVWVSYLASISFSLWVRIDPFSRYGPDILE 1060
 DB 1015 FFAFWIVHLVLPFLKGLMGHONRPTIVVWVSYLASISFSLWVRIDPFSRYGPDILE 1074
 QY 1061 CGINC 1065
 DB 1075 CGINC 1079

RESULT 9

AAV84121 standard; protein; 1076 AA.

AAV84121;

03-JUL-2000 (first entry)

DNA encoding a maize cellulose synthase.

Maize; cellulose synthase; stalk quality; stand; silage; cellulose;

transgenic plant; plant breeding marker.

Zea mays.

Key Location/Qualifiers

Misc-difference 425 /note="encoded by ANG"

W0200009706-A2.

16-AUG-1999; 99WO-US18760.

17-AUG-1998; 98US-0096822.

(PION-) PIONEER HI-BRED INT INC.

Dhuga KS, Helentjaris TG, Bowen BA, Wang X;

WPI; 2000-224343/19.

N-PADB; AAZ99533.

New genes which encode maize cellulose synthase polypeptides in plants

useful for modulating the expression of cellulose synthase in plants

and to produce transgenic plants expressing the novel protein

Claim 15; Page 196-199; 119pp; English.

The present sequence represents a maize cellulose synthase polypeptide.

The cellulose synthase can be used for the improvement of stalk quality

of cellulose in the pericarp, hardening the kernel and improving its

CC handling ability. The sequences are used to produce transgenic plants
 CC and seeds expressing the cellulose synthase. The polynucleotide is
 CC used for modulating, preferably increasing, the level of the synthase
 CC in a plant cell. The plants are preferably monocots. The polynucleotide
 CC is also used as a probe or primer in the detection quantitation or
 CC isolation of gene transcripts. The probes are useful in detecting
 CC deficiencies in the level of mRNA in screenings for desired transgenic
 CC plant, for detecting mutations in the gene, for monitoring upregulation
 CC of expression or changes in enzyme activity in screening assays of
 CC compounds, for detection of any number of allelic variants of the gene,
 CC or for use as molecular markers in plant breeding programs. The
 CC isolated nucleic acids of the present invention can also be used for
 CC recombinant expression of their encoded polypeptides or for use as
 CC immunogens in the preparation and/or screening of antibodies. The
 CC proteins can be employed in assays for enzyme agonists or antagonists
 CC of enzyme function or for use of immunogens or antigens to obtain
 CC antibodies specifically immunoreactive with a protein.

Sequence 1076 AA:

Query Match 79.8%; Score 4529.5; DB 21; Length 1076;

Best Local Similarity 78.1%; Pred. No. 0;

Matches 847; Conservative 88; Mismatches 118; Indels 31; Gaps 10;

QY 4 EGSETAGKPMKNIVPOTQICSDNVGKTVDGDFVADICSPVCRPCYERKGNQSCP 63

DB 2 DGGDATNSGKHVAGVQICDGDGVGAADGDLTACDVCFCPCYERKEDGTQACP 61

QY 64 QCKTKYKRLKGSFAIRGDEDEGLADEGVFEFVY-----POKERTISEMIGLHLRGKG 117

DB 62 QCKTKYKRLKGSFAIRGDEDEGLADEGVFEFVY-----POKERTISEMIGLHLRGKG 119

QY 118 EEMGEPOYDK-----EYSHNLPLRITSROPTSGEFSASAPERT--SVSYTIAGG 164

DB 120 SDIGLAKYDSGELGCHKYDSGELPRGYITSLHSQ--ISELIPGASPDHMSPPGNIGRGG 178

QY 165 KRLPYSSDVNQSFRRIYDPVIGVNAKERYDGMKQOE--KNTGPVSTQASERGV 221

DB 179 HGPFY---VNHSPNPREFSGSLGNVAKERYDGMKQCAIPMTNGTSIAPSEGRGVA 235

QY 222 DIDASTDLADEALNDEAROPLSRVSPSSINRYAVIMRLVITLFLHYRTINPV 281

DB 236 DIDASTDYNNEDLNDERTROPLSRVSPSSINRYAVIMRLVITLFLHYRTINPV 295

QY 282 PNAFALMLVSVCEIWFALSMIIDPPKMPVRYRETYLRLALRYDREGEPSQLAAVDIF 341

DB 296 NNAIYPLMLLSVCEIWFALSMIIDPPKMPVRYRETYLRLALRYDREGEPSQLAAVDIF 355

QY 342 VSTVDLPKEPPLVYANTVLSILAVDYPVDKVCYVFDGGAAMLSEFSLAETSEFARKWP 401

DB 356 VSTVDLPKEPPLVYANTVLSILAVDYPVDKVCYVFDGGAAMLSEFSLAETSEFARKWP 415

QY 402 FCKKYSIEPRAPRWYAAKIDYIKDKVQTSFYKDRAMREVEEFIRINALYSKALRKP 461

DB 416 FCKKYNIEPRAPRWYAAKIDYIKDKVQTSFYKDRAMREVEEFIRINALYSKALRKP 475

QY 462 EEEGVNODCTPMPGNTGHPGMIQVFLONGGLDEGNEELRLVYVSHREKRPGEFHKK 521

DB 476 EEEGVNODCTPMPGNTGHPGMIQVFLONGGLDEGNEELRLVYVSHREKRPGEFHKK 535

QY 536 AGAMNALVRSVAVLTNGOYMLNDCDHYINNSKALEAMCFIAMDPLGRNVCYVOPRF 595

DB 582 DGDIDKDRANRNTVFQDNLGIDGIDGPPVVGCCVNNRATLGYEPRIVKHKRPSL 641

QY 596 DGDIDKDRANRNTVFQDNLGIDGIDGPPVVGCCVNNRATLGYEPRIVKHKRPSL 653

QY 642 LSKLGGSRKKNSKAKESDKKSGHRTSTYVFNLDIDIEGVEGAGFDEKALMSQM 701

DB 654 FSLICGG-KKKTSKSKSEKKSHRHADSSVPVFLNDEBEGISGQPDDEKSLMSQM 712

QY 702 SLEKRGASAVVASTLMENGVPSPATPENLKEAIVHISCGYEDKSDMGMEIGMIYGS 761

CC	Sequence	1075 AA:	71.0%;	Score 4028.5;	DB 21;	Length 1075;
CC	isolated nucleic acids of the present invention can also be used for					
CC	recombinant expression of their encoded polypeptides or for use as					
CC	immunogens in the preparation and/or screening of antibodies. The					
CC	proteins can be employed in assays for enzyme agonists or antagonists					
CC	of enzyme function or for use of immunogens or antigens to obtain					
CC	antibodies specifically immunoreactive with a protein.					
XX						
SO	Sequence	1075 AA:				
	Query Match		71.0%;	Score 4028.5;	DB 21;	Length 1075;
	Best Local Similarity		69.6%;	Pred. No. 0;		
	Matches	752;	Conservative 130;	Mismatches 153;	Indels 45;	Gaps 12
OY	4 EGETAG--KPKMNIYPTOTCISDNVYKTVGDGRFVACDSCFPRPCRYEYERKDGNS	61				
Db	23 DGDVPGSAKPTKSNAGQYCOICGDSVGSATGDVFAVACNECAFVPCRPCEYERKGNOC	82				
OY	62 CPOCKTRKRLKSGAIPGDDDEDLDEGTYEFNYPQEKASERLGNHLRGGEENG	121				
Db	83 CPOCKTRKRLKSGSPRVHGEDEDEY--DLDLNEFNKQGS-----GKPEMQ	128				
OY	122 EPQYDKEYS-----NHMLPRLTSRODTSGEFSAAPERLVSSTIAGKRLPYSDPN	174				
Db	129 LQGDADADLSSARHPEHRIPLRTSGQOISGRIPLASPRHRI-----KSPISSTYD	180				
OY	175 QS--PNRIIVDP-----VGLGNVAMKERVDGMMKKOEKNTGVPVTOASERGVDDIAS	226				
Db	181 PSVPAPVPIYDPSKDLNSYGLNSVDMKEVESMRVAKODNMQVTKRYPEARGG--DME--G	238				
OY	227 TDILADELLDDEAQRPLSRKSYIFSSRTNPRYMTIMRLVYLICFLYHRIINPRVNAFA	286				
Db	239 TGSNGEXMOMDDAALPLSRIPYISSNOLNLRVYIILRLILCFEFGYRVSHPRVDAVG	298				
OY	287 LMLVSVCEIMFALSMLIDQEPKMPVPVNETYLDRLALRYDREGSPQAAVDIEFSTVD	346				
Db	299 LMLVSVCEVWFALSMLIDQEPKMPVPIRETILDLRLALRYDREGSPQALPDIYFSTVD	358				
OY	347 PLKEBPVLTANTVLSILAADVVDKVCSCVFPDGAAMLSESLAETSEBARMKVPRCKKY	406				
Db	359 PLKEBPVLTANTVLSILSVDPVDKVCSCVSDGSAMLTFESLSETAEBARMKVPRCKKH	418				
OY	407 SIERRAPMYTAAKIDLYLKDQVQTSFYVDRAKMAKREYEFKIRINALVSKALCPREGNV	466				
Db	419 NIEPRAPPEFYAOKIDYLLKDKIQPSFVERRAKMYREYEFKIRINALVAKOQVPRREGNT	478				
OY	467 MODGPRWGNNTMGHRCGMIOYFLONGGLDAGCNELPRLYUVYSREKREPOHNKKAGANN	526				
Db	479 MADGPRWGNNTMRDHPGMIOYFLHSGSGLDTDGNEPLRYUVYSREKREPOHNKKAGANN	538				
OY	527 ALVRYSAVLTPGPFILNLDCHYIINNSKALREAMCFGLMDPNLGKQVCYQVQFORFGIDK	586				
Db	539 ALIRYSAVLTPGAYILNVDCHYIFNSSKALREAMCFPMDRPALGRKVCYQVQFORFGIDL	598				
OY	587 NDRYANRNTVFEDINLRGLDGIQGFVYVGTGCVFNRTALYXEPRIKVAHKRPSLSKLC	646				
Db	599 HDRYANRNTVFEDINMKGLDGIQGFVYVGTGCFNROALYXGDPVLTEADLRPNYIKSC	658				
OY	647 GGSRRKNSKAKKESKKSGRTSDTVFVFNLDIEEGVEGAGFDFDEKALLMSQNSLEKR	706				
Db	659 CGRRKKKKKSYMDSQSRIMKR--TESSAPFIENNEDEIEEGE--GYEDERSVLSORLEKR	715				
OY	707 FGQSAVFASLTMEENGVPSPATPNLLKLEALHIVISCGEDKSDMCMEIGMTIYGSUTEDI	766				
Db	716 FGQSPFIFASTFMTGGIPTSPINPASPALKLEALHIVISCGEDKTEMKELGWTIYGSUTEDI	775				
OY	767 LTGFRMARHGRNSIYCMKPLPAFKGSAPIINSDRIQVLRMALGSEYIELSRHRCPIYGY	826				
Db	776 LTGFEMHARGMOSIYCMPPRCFKGSAPINISDRINQVLRMALGSEYIELSRHRCPIYGY	835				
OY	827 NGRKLFLERAVVNTTIYPTISIPILMCTLLAVCLFTNQFIITPQISNASTIWFSLFSLS	886				
Db	836 NGRKLFLERAVIINTIYPTISIPILACVLPALICLTNNKFIITPISNAGAFILLFAS	895				

DB 956 DEDGFAELVFKMTSLIPPTVIVINLVGVAGISVAINSGYSGMGLFEGKLFPSIMV 1015
QY 1007 IVHLVPLFKGLMGROKRTPTIYVWMSVLLASIFSLMWRIDPFTSRVTPDL-ECGINC 1065
DB 1016 ILHLVPLFKGLMGROKRTPTIYVWMSVLLASIFSLMWRIDPFTSRVTPDL-ECGINC 1075
RESULT 14
ID AAY84110
AAY84110 standard; Protein; 1074 AA.
AC AAY84110;
DF 03-JUL-2000 (first entry)
DE Amino acid sequence of a maize cellulose synthase.
KM Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
transgenic plant; plant breeding marker.
XX Zea mays.
XX WO200009706-A2.
XX 24-FEB-2000.
XX 16-AUG-1999; 99WO-US18760.
XX 17-AUG-1998; 98US-0096822.
PA (PION-) PIONEER HI-BRED INT INC.
PI Dhugga KS, Helentjaris TG, Bowen BA, Wang X;
DR WPI: 2000-224343/19.
DR N-PSDB: AA299500.
PT New genes which encode maize cellulose synthase polypeptides in plants
PT useful for modulating the expression of cellulose synthase in plants
PT and to produce transgenic plants expressing the novel protein -
PS Claim 15; Page 110-113; 119pp; English.
XX The present sequence represents a maize cellulose synthase polypeptide.
CC for improved stand or silage. It also provides an increased concentration
CC of cellulose in the pericarp, hardening the kernel and improving its
CC handling ability. The sequences are used to produce transgenic plants
CC and seeds expressing the cellulose synthase. The polynucleotide is
CC used for modulating, preferably increasing, the level of the synthase
CC in a plant cell. The plants are preferably monocots. The polynucleotide
CC is also used as a probe or primer in the detection quantitation or
CC isolation of gene transcripts. The probes are useful in detecting
CC deficiencies in the level of mRNA in screenings for desired transgenic
CC plant, for detecting mutations in the gene, for monitoring upregulation
CC of expression or changes in enzyme activity in screening assays of
CC compounds, for detection of any number of allelic variants of the gene,
CC or for use as molecular markers in plant breeding programs. The
CC isolated nucleic acids of the present invention can also be used for
CC recombinant expression of their encoded polypeptides or for use as
CC immunogens in the preparation and/or screening of antibodies. The
CC proteins can be employed in assays for enzyme agonists or antagonists
CC of enzyme function or for use of immunogens or antigens to obtain
CC antibodies specifically immunoreactive with a protein.
SQ Sequence 1074 AA;
Query Match 70.9%; Score 4024; DB 21; Length 1074;
Best Local Similarity 69.4%; Pred No. 0;
Matches 749; Conservative 128; Mismatches 159; Indels 44; Gaps 10;

DB 23 DGDADVPAKPTSRANSAGQVQOIGDPTVGVASATGDVFAVACNECAFPCRCPEYERKRGNC 82
QY 62 CPOCKTRKRLKGSFPAIPEDCKDEGLADEGYEFNTPQKEKISERMLGHHLTRGGEENG 121
DB 83 CPOCKTRKRLKGSFPAIPEDCKDEGLADEGYEFNTPQKEKISERMLGHHLTRGGEENG 128
QY 122 EPQYDKEYS-----HNHLPRLTSSKODTSGEFSAASPERLSVSSRTIAGCKRLPYSSDVN 174
DB 129 LOGDDADSSSARHDPHRIPLRTSGQOISGETPDASPDHRSI-----RSPSTSYVD 180
QY 175 QS--PNRRIYDP-----VGLGNVAMKERVDCKMKMKOEKNTGPSTOAASEGGVDIDAS 226
DB 181 PSVPVPRIVDPSKDLNSGLNSGVDMKEVESMRVKODKNMLQVNTKPYEARGDME--G 237
QY 227 TDILADEALLDEANOPLSRKVSITSSSRINPIMVIMLVLCLFLHRTNPPVNAFA 286
DB 238 TGSNGEDMQVMDARPLSRIVPISSNQNLRIYIILRLIILCEFFOYRISHPYRNAYG 297
QY 287 LMLVSVICWIFALSMIILDOFPKMPVNRETYLDRLALRYDREGPSQLAPIDVFSTVD 346
DB 298 LMLVSVICWIFALSMIILDOFPKMPVNRETYLDRLALRYDREGPSQLAPIDVFSTVD 357
QY 347 PLKEPPLVANTVLSILAVIDPVDRVSCYVPDGAAMLSESLAETSEFARKVPECKKY 406
DB 358 PLKEPPLVANTVLSILAVIDPVDRVSCYVPDGAAMLSESLAETSEFARKVPECKKY 417
QY 407 STEPPAPEWYFAAKIDYLDKQVQTSFVKDRAMKREYEFKTRINALYSKALCKPEEGV 466
DB 418 NIEPRAPEEYFAOKIDYLDKQVQTSFVKDRAMKREYEFKTRINALYSKALCKPEEGV 477
QY 467 MODGTPWGNNTGDHPGMOVFLGONGGLDAGNENLPRLYVYSREKRPFGOHKKAGAN 526
DB 478 MADGTAMPENNRDRHGMIOVFLGONGGLDAGNENLPRLYVYSREKRPFGOHKKAGAN 537
QY 527 ALVRSAYVLTNGPFLTNDGDHYINNSKALREAMCFMLDPNGLKQVYQVOPORFIDDK 586
DB 538 ALVRSAYVLTNGPFLTNDGDHYINNSKALREAMCFMLDPNGLKQVYQVOPORFIDDK 597
QY 587 NDRYANRNTVFEDINLRGIDIGQGYVYGTGCVFNRTALYGEPTPKVHKKPSSLKSC 646
DB 598 HDRYANRNTVFEDINLRGIDIGQGYVYGTGCVFNRTALYGEPTPKVHKKPSSLKSC 657
QY 647 GGRKRSKAKKESKSKKSGRHTDSTVPFENLDIDEEGEGAFDDEKALLKNSOMSLER 706
DB 658 GGRKRSKAKKESKSKKSGRHTDSTVPFENLDIDEEGEGAFDDEKALLKNSOMSLER 714
QY 707 FGOSAVFVATLMENGVPSPATPENLKEAIVHISCGYEDKSDMGMEIGWYGVSTEDI 766
DB 715 FGOSAVFVATLMENGVPSPATPENLKEAIVHISCGYEDKSDMGMEIGWYGVSTEDI 774
QY 767 LTFGMHARGMSIYCMPLPAFKGSAPIINLSDRLQVLRALGSEVILFSRHCPTWY 826
DB 775 LTFGMHARGMSIYCMPLPAFKGSAPIINLSDRLQVLRALGSEVILFSRHCPTWY 834
QY 827 NGRLEFLERFAVNTTYITISIPILMYCTLLAVCLFTNQFIIPQSNIASIWFSLFSL 886
DB 835 NGRLEFLERFAVNTTYITISIPILMYCTLLAVCLFTNQFIIPQSNIASIWFSLFSL 894
QY 887 IFATGILERMWSGVGIDEMWRNDEQFVIGTSAHLFAVPOGLLKVLAGIDTNPYTSKAS 946
DB 895 IFATGILERMWSGVGIDEMWRNDEQFVIGTSAHLFAVPOGLLKVLAGIDTNPYTSKAS 954
QY 947 DEDGFAELVFKMTSLIPPTVIVINLVGVAGISVAINSGYSGMGLFEGKLFPSIMV 1006
DB 955 DEDGFAELVFKMTSLIPPTVIVINLVGVAGISVAINSGYSGMGLFEGKLFPSIMV 1014
QY 1007 IVHLVPLFKGLMGROKRTPTIYVWMSVLLASIFSLMWRIDPFTSRVTPDL-ECGINC 1065
DB 1015 ILHLVPLFKGLMGROKRTPTIYVWMSVLLASIFSLMWRIDPFTSRVTPDL-ECGINC 1074
RESULT 15

AA84112
ID AA84112 standard; Protein; 1074 AA.
XX
AC AA84112;
XX
DT 03-JUL-2000 (first entry)
XX
DE Amino acid sequence of a maize cellulose synthase.
XX
KW Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
transgenic plant; plant breeding marker.
XX
OS Zea mays.
XX
PN WO200009706-A2.
XX
PD 24-FEB-2000.
XX
PE 16-AUG-1999; 99MO-US18760.
XX
PR 17-AUG-1998; 98US-0096822.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Dhuga KS, Helentjaris TG, Bowen BA, Wang X;
XX
DR WPI: 2000-224343/19.
XX
DR N-PSDB; AA299506.
XX
PT New genes which encode maize cellulose synthase polypeptides in plants
PT useful for modulating the expression of cellulose synthase in plants
PT and to produce transgenic plants expressing the novel protein
PS Claim 15; Page 126-128; 119pp; English.

XX The present sequence represents a maize cellulose synthase polypeptide.
XX The cellulose synthase can be used for the improvement of stalk quality
XX for improved stand or silage. It also provides an increased concentration
XX of cellulose in the pericarp, hardening the kernel and improving its
XX handling ability. The sequences are used to produce transgenic plants
XX and seeds expressing the cellulose synthase. The polynucleotide is
XX used for modulating, preferably increasing, the level of the synthase
XX in a plant cell. The plants are preferably monocots. The polynucleotide
XX is also used as a probe or primer in the detection quantitation or
XX isolation of gene transcripts. The probes are useful in detecting
XX deficiencies in the level of mRNA in screenings for desired transgenic
XX plants, for detecting mutations in the gene, for monitoring upregulation
XX of expression or changes in enzyme activity in screening assays of
XX compounds, for detection of any number of allelic variants of the gene,
XX or for use as molecular markers in plant breeding programs. The
XX isolated nucleic acids of the present invention can also be used for
XX recombinant expression of their encoded polypeptides or for use as
XX immunogens in the preparation and/or screening of antibodies. The
XX proteins can be employed in assays for enzyme agonists or antagonists
XX of enzyme function or for use of immunogens or antigens to obtain
XX antibodies specifically immunoreactive with a protein.

Sequence 1074 AA;

Query Match 70.9%; Score 4024; DB 21; Length 1074;
Best Local Similarity 69.4%; Pred. No. 0;
Matches 749; Conservative 128; Mismatches 159; Indels 44; Gaps 10;

QY 2 ESEGETAGKPMKNIYPTQICSDNYGKTVDDRFVACIGCSFPCYCEYERKGNOS 61
DB 23 DDDAPVPAKPTKSANQVOICGDIYVGSATGDFVACBACFPVPCPEYERKGNOC 82
QY 62 CPQCKTRRYRLKGLPAIPDKDGLADEGTFVFNYPQEKISERMLGWHLTGKGEENG 121
DB 83 CPQCKTRRYRKQKSPRVHGDDEEDY-DLLDNEFNKYK-----GNGKGPENQ 128
QY 122 EPOYKREVS-----HNLPLRLTSKQDTSGEFSNAPSRLSVSTIAGCKRLPYSDDV 174

DB 129 LQGDADLSSSARHDPHRIPLRTSGQISGETIPDASPDNRHSI-----NSPISSYYD 180
QY 175 QS--PNNRIYDP-----VGLGNVANKERVDGKMKOEKNTGPGVSTQASERGVDIDAS 226
DB 181 PSVPVPRVRIYDPSKDLNSYGLNSVDMKEKVESRNVQDMNLQYTKRYEAGDMR---G 237
QY 227 TDILADEALLNDAROPLSRKVSIPSSRIYPRVMTLRVLVILCEFLHYRTINPVNARA 286
DB 238 TSGNGEDMQVDARPLSRIVPSSNQLNRYIYILRLILCEFFQYRISHPVNAYG 297
QY 287 LMLVSYICITWPLSMIILOFPKMFVFNRETYIDRLALYVDGEGESQLAADIFPSTD 346
DB 298 LMLVSYICVWFALSMIILOFPKMFVFNRETYIDRLALYVDGEGESQLAADIFPSTD 357
QY 347 PLKEPPLTANTVPLSTLAVDYVDKVCYVFPDGAAMLSPESLAETSEFARKVPPCKRY 406
DB 358 PLKEPPLTANTVPLSTLAVDYVDKVCYVFPDGAAMLSPESLAETSEFARKVPPCKRY 417
QY 407 STEPRAPWYFAKIDYLDKQVTSFVKDRAMKREYEFKIRINALVSKALCPREGVY 466
DB 418 NIEPRAPWYFAKIDYLDKQVTSFVKDRAMKREYEFKIRINALVSKALCPREGVY 477
QY 467 MDQGTWPGNNTGDHGMIOVFLGONGIDABENELPRVLYYSREKRPFGHKKRGAAN 526
DB 478 MADGTAMPGNNDHGMIOVFLGONGIDABENELPRVLYYSREKRPFGHKKRGAAN 537
QY 527 ALVRSVAVLTNGPFIINTLDCDHYINNSKALREAMECLMPDNLGQCYVQFOPFGIDK 586
DB 538 ALVRSVAVLTNGPFIINTLDCDHYINNSKALREAMECLMPDNLGQCYVQFOPFGIDK 597
QY 587 NDYRANRNTVPEDINLRGIDGIDGYPVYVGTGCVFNFTALYGEPIKVRHKKPSLSKIC 646
DB 598 HRYRANRNTVPEDINLRGIDGIDGYPVYVGTGCVFNFTALYGEPIKVRHKKPSLSKIC 657
QY 647 GGRKKNKAKKESDKKKKSGRHTDVPVFNDDIEEGVAGFODEKALNOMSLER 706
DB 658 GGRKKNKAKKESDKKKKSGRHTDVPVFNDDIEEGVAGFODEKALNOMSLER 714
QY 707 FGQSAFVAVSTLMENGVPPSATPENILKEALHIVISGGEYEDKSDMGMEIGWYGVTEDI 766
DB 715 FGQSPFIATSTFTMGQIGIPSTNPASLKEALHIVISGGEYEDKSDMGMEIGWYGVTEDI 774
QY 767 LTGFKHANGWNSIYCMPLPAFKGAPINLSDRNLQVLRMALGSVEILLSHRCPWGY 826
DB 775 LTGFKHANGWNSIYCMPLPAFKGAPINLSDRNLQVLRMALGSVEILLSHRCPWGY 834
QY 827 NGRLKLEPRFAYVNTIYPTISPILMYCTLLAVCLFTNOPIIPOSINASTIMPLSTFIS 886
DB 835 NGRLKLEPRFAYVNTIYPTISPILMYCTLLAVCLFTNOPIIPOSINASTIMPLSTFIS 894
QY 887 IFATGILEMRWSGVGIDEMWRNDEQFVIGVSAHLFAVQGLIKVLADITNFTVYSKAS 946
DB 895 IFATGILEMRWSGVGIDEMWRNDEQFVIGVSAHLFAVQGLIKVLADITNFTVYSKAS 954
QY 947 DEGDGFADLYLFRKWTLLIPPTTLVLNVLGVAVGVSAIVNSGYSGWGLFGKLFPAFV 1006
DB 955 DEGDGFADLYLFRKWTLLIPPTTLVLNVLGVAVGVSAIVNSGYSGWGLFGKLFPAFV 1014
QY 1007 IYHLVPELKGMLGRQKRPETIVVMSVLASTFSLMLVITDPTSVTPDIL-DEGINC 1065
DB 1015 IYHLVPELKGMLGRQKRPETIVVMSVLASTFSLMLVITDPTSVTPDIL-DEGINC 1074

Search completed: June 16, 2003, 11:06:41
Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:03:13 ; Search time 15 Seconds

(without alignments)
2944.819 Million cell updates/sec

Title: AAC39336

5677

Sequence: 1 MESEGETAGKPMKNIVPQC.....IDPFTSRVGPDIIECGINC 1065

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	273	4.8	872	1	BCSA_ECO57
2	273	4.8	874	1	BCSA_SALTI
3	272	4.8	872	1	BCSA_ECOLI
4	271	4.8	874	1	BCSA_SALTY
5	253.5	4.4	729	1	BCSA_XANAC
6	253.5	4.4	1596	1	ACS2_ACCEX
7	247.5	4.4	1518	1	BCA4_ACCEX
8	236.5	4.4	1518	1	BCA5_ACCEX
9	236.5	4.2	756	1	BCA2_ACCEX
10	228.5	4.0	754	1	BCA1_ACCEX
11	228	4.0	739	1	BCSA_PSEFL
12	208.5	3.7	745	1	BCA3_PSEFL
13	196	3.5	1550	1	ACS1_ACCEX
14	119	2.1	517	1	COX1_PARLI
15	118.5	2.1	590	1	YKTA_CAEBL
16	115	2.0	357	1	G6PT_HUMAN
17	115	2.0	517	1	COX1_ASTPE
18	110.5	1.9	441	1	YCDQ_ECOLI
19	110.5	1.9	672	1	ALIB_HUMAN
20	110	1.9	517	1	COX1_STRPU
21	109	1.9	830	1	FAR1_YEAST
22	108	1.9	517	1	COX1_PISOC
23	108	1.9	1029	1	RIP3_RAT
24	107.5	1.9	1437	1	MRP5_HUMAN
25	107	1.9	3033	1	POLG_HCV6
26	106	1.9	357	1	G6PT_CANFA
27	105.5	1.9	1076	1	YE38_SCHPO
28	105.5	1.9	4829	1	BIR6_HUMAN
29	104.5	1.8	309	1	YESP_BACSU
30	104	1.8	995	1	AHM7_ARATH
31	104	1.8	1167	1	CIAA_BACTU
32	104	1.8	1879	1	POLN_SMSV1
33	103	1.8	414	1	YOAB_BACSU

34	103	1.8	470	1	ROCC_BACSU
35	103	1.8	1066	1	NUC2_NEUCR
36	103	1.8	1232	1	B3A3_HUMAN
37	102	1.8	642	1	ARE2_YEAST
38	101.5	1.8	1863	1	BRC1_HUMAN
39	101	1.8	617	1	VAA1_BOVIN
40	101	1.8	3097	1	CADN_DROME
41	100.5	1.8	744	1	YVW2_CAEBL
42	100.5	1.8	1020	1	ATC1_DROME
43	99.5	1.8	322	1	NTM_STRPU
44	99.5	1.8	536	1	COX1_ALUMA
45	99.5	1.8	714	1	HUNK_MOUSE

ALIGNMENTS

RESULT 1
BCSA_ECO57
ID BCSA_ECO57 STANDARD: PRT: 872 AA.
AC 08XSL7:
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).
GN BCSA OR 24948 OR EC54413.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDE933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Evans P.S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamosis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RT Nature 409:529-533(2001).
[2]
SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIND 0509952;
RX MEDLINE=21156231; PubMed=11238796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RT DNA Res. 8:11-22(2001).
-1- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes
uridine 5'-diphosphate glucose to cellulose, which is produced as
an extracellular component for mechanical and chemical protection
at the onset of the stationary phase, when the cells exhibit
multicellular behavior (rod morphology). Co-expression of
cellulose and thin aggregative fimbriae leads to a hydrophobic
network with tightly packed cells embedded in a highly inert
matrix (by similarity).
-1- CATALYTIC ACTIVITY: UDP-glucose + ((1,4-beta-D-glucosyl))_(N) = UDP
+ ((1,4-beta-D-glucosyl))_(N+1).
-1- CORRECTOR: Magnesium (by similarity).
-1- ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid
(C-di-GMP) (by similarity).
-1- PATHWAY: Bacterial cellulose biosynthesis.
-1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
(potential).
-1- DOMAIN: There are two conserved domains in the globular part of
the protein: the N-terminal domain (domain A) contains
the conserved DXD motif and is possibly involved in catalysis and
substrate binding. The C-terminal domain (domain B) contains the

-1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

EMBL; AL672781; CAD08006.1;
DR InterPro; IPR001173; Glycosyltransferase_2.
DR Pfam; PF00535; Glycos_transf.2; 1.
KW Cellulose biosynthesis; Transferase; Glycosyltransferase;
KW Transmembrane; Inner membrane; Complete proteome.
FT DOMAIN 271 364
FT TRANSMEM 441 501 CATALYTIC SUBDOMAIN A.
FT TRANSMEM 30 50 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 173 193 POTENTIAL.
FT TRANSMEM 230 250 POTENTIAL.
FT TRANSMEM 525 545 POTENTIAL.
FT TRANSMEM 547 567 POTENTIAL.
FT TRANSMEM 592 612 POTENTIAL.
FT TRANSMEM 634 654 POTENTIAL.
FT TRANSMEM 668 688 POTENTIAL.
FT TRANSMEM 833 853 POTENTIAL.
FT ACT_SITE 313 313 POTENTIAL.
FT SITE 457 457 POTENTIAL.
FT SITE 360 360 SUBSTRATE BINDING (POTENTIAL).
FT SITE 362 362 SUBSTRATE BINDING (POTENTIAL).
SQ SEQUENCE 874 AA; 100020 MW; 3395D09CD0F51990 CRC64;

Query Match Best Local Similarity 4.8%; Score 273; DB 1; Length 874;
Matches 153; Conservative 91; Mismatches 201; Indels 370; Gaps 29;

DQ 260 MVIVLRVVLICLFHYRTNV-----PAKFLMVSVCETLFWLSWLLDDPPKKFPYNR 315
Db 201 MLVSLSTSVSCRYIWMRTSTLNMDDPSLVCGILLFAEYYAMVAVLGFOVVAPLNIR 260
DY 316 ETLYDLRLATRDREGEPSOLAOFVETVDYORLEKPPLVTNTVLTSLAIVDPDVACY 375
Db 261 QP-----VPLEKNSQMPTVIDIEFTYN---EDLNVAKNITIASGIDMPCKKLNIW 309
DY 376 VFDDGAAMLSESIASLETSEFAKKWVPESCKYSIERBAEWYFAAKIDYLDKVQTSFYKD 435
Db 310 ILDDGR---ESP--RQFAR----- 324
DY 436 RRANKREYEERKITINALVALSKCRREGGWMOGSTPWPGNNTGDHGMIOVFAGNGGL 495
Db 325 -----HVGN----- 329
DY 496 DAEGNELRLUYVSREKRPFQHNKKAGAMALRVSAVLNLGPRIIMLDCDHYNINSKA 555
Db 330 -----YLR---ATHNAKAGINNALNHA---KEEVALLFDCHDVTRSF 370
DY 556 LREAMSGLMDFNLKOVUCVOPORF---DEIDKN---DRYANENTVFFDIINLKGDIGI 608
Db 371 QMTGWMLFE----KOLAMQTPIHFERRPDREFERNLGSFRKTRPNNGTLFGIGUDGNDMM 426
DY 609 OGPRUUVGTGCYFNFTALTYGUERPPIKVKHKPSLLSKLCGGSSKKKSAAKSDKKKSGRH 668
Db 427 DATEFCSSCAVIRR-----KP----- 442
DY 669 TDSIVPVAFNLDIDEGVGAGFDEDKALLMSOMSLERKFQGSAVAVASTIMEGVPPSA 728
Db 443 ----LDKI-----GGI----- 449

QY	729	TPENLLKEAHIVNICGVEDKSDMGMEIGWICSTEDILITGFKMHARGNRSTYCMKPKPA	788
Db	450	-----AAE-----: : : : : : :	478
QY	789	EKSGAPINLSORLNQVLRMALGVSVEILFESRCPITWGYNGRLKELERFAVNTTTPITS	848
Db	479	SAGLATESLSAHIGQRIIRMAKGWQI-FRLDNPL---FGKGLKLAORLCYLNAMFHLSS	534
QY	849	IPLLMTCY-----LLAVCLTNNQFIIPQI-----SNIASINPLSFLSFA	889
Db	535	IPRLIFLTAPLAFLLLAHYIYAPALMALEFVIFHWYHASTLNKRGKYSRHSWSEIYE	594
QY	890	TGILEMRMSGVIGIDEMWRNEQFVIGVSAAHLFAVFOGLKVLKVLADTNNFYFESKASDD	949
Db	595	TVLA-----WTIAPPT-----LVALINPHKGFENYK-----	622
QY	950	GDFAEALYLFKWTLLIPPTLLIYNLVGVAGVSAYAINSGVSMPLFELFFAFWIVH	1009
Db	623	GGTVEEKYVDM-VISRYIFLVLNLLGVAAGV-----WRYVYCP-----	661
QY	1010	LYPLKGLMGKRONPTPIVWVWSVLASIFSLLAV	1044
Db	662	-----ENETLTVIV-----SLVWV	675
RESULT 3			
BCSA.ECOLI			
ID	BCSA.ECOLI	STANDARD:	PRF: 872 AA.
AC	P37653; P37654; P76712; P76713;		
DT	01-OCT-1994 (Rel. 30, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Cellulose synthase catalytic subunit [BDP-forming] (EC 2.4.1.12).		
GN	BCSA OR B3553.		
OS	Escherichia coli.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Escherichia.		
OX	NCBI_TaxID=562;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-K12 / MG1655.		
RX	MEDLINE=94316500; Pubmed=8041620;		
RA	Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;		
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,		
RA	Gregor J., Davis N.W., Klinkpatrick H.A., Goeden M.A., Rose D.J.,		
RA	Mau B., Siao Y.;		
RT	"The complete genome sequence of Escherichia coli K-12.";		
RL	Science 277:1453-1474(1997).		
RN	[3]		
RP	CHARACTERIZATION		
RC	STRAIN=ECOR10, ECOR12, and TOB1;		
RX	PubMed=11260463;		
RA	Zogar X., Niantz M., Rohde M., Bokranz W., Roemling U.;		
RT	"The multicellular morphotypes of Salmonella typhimurium and		
RT	Escherichia coli produce cellulose as the second component of the		
RT	extracellular matrix.";		
RL	Mol. Microbiol. 39:1452-1463(2001).		
CC	-1- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes		
CC	uridine 5'-diphosphate glucose to cellulose, which is produced as		
CC	an extracellular component for mechanical and chemical protection		
CC	at the onset of the stationary phase, when the cells exhibit		
CC	multicellular behavior (rdar morphotype). Co-expression of		
CC	cellulose and thin aggregative fimbriae leads to a hydrophobic		
CC	network with tightly packed cells embedded in a highly inert		

matrix.
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + ((1,4-beta-D-glucosyl))(N) = UDP
 CC + ((1,4-beta-D-glucosyl))(N+1).
 CC -1- COFACTOR: Magnesium (By similarity).
 CC -1- ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid
 CC (c-di-GMP).
 CC -1- PATHWAY: Bacterial cellulose biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (potential).
 CC -1- DOMAIN: There are two conserved domains in the globular part of
 CC the protein: the N-terminal domain (domain A) contains the
 CC conserved DXD motif and is possibly involved in catalysis and
 CC substrate binding. The C-terminal domain (domain B) contains the
 CC QXXRW motif and is present only in processive glycosyl
 CC transferases. It could be involved in the processivity function of
 CC the enzyme, possibly required for holding the growing glycan chain
 CC in the active site.
 CC -1- MISCELLANEOUS: The genes bscA, bscB, bscZ and bscC are
 CC constitutively transcribed but cellulose synthesis occurs only
 CC when adaA, a putative transmembrane protein regulated by agfD, is
 CC expressed. Cellulose production is abolished in E.coli K12.
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to a
 CC frameshift in position 128.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U00039; AAB18510.1; ALT_FRAME.
 CC EMBL: U00039; AAB18511.1; ALT_FRAME.
 CC EMBL: A000430; AAC76558.1; ALT_INIT.
 CC DR Ecogene: E012260; bcsA.
 CC DR InterPro: IPR001173; Glycosyltransf_2.
 CC DR Pfam: PF00535; Glycosyltransf_2; 1.
 CC KM Cellulose biosynthesis; Transferase; Glycosyltransferase;
 CC Membrane; Inner membrane; Complete proteome.
 CC FT DOMAIN 271 364 CATALYTIC SUBDOMAIN A.
 CC FT 441 501 CATALYTIC SUBDOMAIN B.
 CC FT TRANSMEM 30 50 POTENTIAL.
 CC FT TRANSMEM 151 171 POTENTIAL.
 CC FT TRANSMEM 173 193 POTENTIAL.
 CC FT TRANSMEM 230 250 POTENTIAL.
 CC FT TRANSMEM 525 545 POTENTIAL.
 CC FT TRANSMEM 547 567 POTENTIAL.
 CC FT TRANSMEM 592 612 POTENTIAL.
 CC FT TRANSMEM 640 660 POTENTIAL.
 CC FT TRANSMEM 668 688 POTENTIAL.
 CC FT TRANSMEM 833 853 POTENTIAL.
 CC FT ACT_SITE 313 313 POTENTIAL.
 CC FT ACT_SITE 457 457 POTENTIAL.
 CC FT SITE 360 360 SUBSTRATE BINDING (POTENTIAL).
 CC FT SITE 362 362 SUBSTRATE BINDING (POTENTIAL).
 CC SO SOURCE 872 AA; 99784 MW; 143268A2EB228F7 CRC64;
 CC -----
 CC Query Match 4.83; Score 272; DB 1; Length 872;
 CC Best Local Similarity 19.08; Pred. No. 1e-11;
 CC Matches 153; Conservative 99; Mismatches 201; Indels 354; Gaps 32;

310 ILDDG----- 315
 436 RRMRKREYEFKIRINALVSKALCKPEEGWMDGTPWPGNNTGDBGMIOVLGONGGL 495
 316 -----REFRQF----- 328
 496 DAGGNELPRLVYSREKRPFGFHHKKAGANALVSAVLTNPFLLNDCDHYINNSKA 555
 329 K-----YARTT-----HEHAKAGINNALKYA-----KGEFVSIFPCDHYPTRSFL 370
 556 LREAMEFLMDPNLKGKQVYQFQRF-----DGIDKN-----DRANRNTVPFDILNRGLDGI 608
 371 QMTGWFLKE-----KOLAMQTHFFSPDPFERNLRGFRKTRNEGTLFGVLQDGDMM 426
 609 QGVVYGTGCVFNRLTALGYEPEPIKVKHKPSLSKLKGGSRKKSKAKESKRRGRH 668
 427 DAFPGSCAVIR-----KP----- 442
 669 TDSVPVFNLDIEEGVEGAGFDEKALMSONSLKRRGQSAVFVASTLMNGVPPSA 728
 443 -----LDEI----- 449
 729 TPENLKEATHVISCGEDEKSDMGMEIGVSTEDILNGFKMHARGRSIDCMKPLA 788
 450 -----AVE-----TVTDAHTSLRHRTGTSAM--RIPO 478
 789 FKSGAPINLSDRLNOVLNMGALSGVELLSHRCPHWYNGR-LKFLERFAYVNTTYPIR 847
 479 AAGLATESLSAHIGORIRMRAGVQI-FRLDNL-----TCKGLKFNORLCYVNAHMFLS 533
 848 SIPLMAYCYLLAVCLFTNOFIPOISNIASIMFLSLFSTFANGILEMRSGVGDENMR 907
 534 GIPRLIFLTPAPLAFLLHAYII-----YAPALMALFV--LPHMHASLSNKGIOGKR 585
 908 NEGFVYIGVSAHLFVAFQGL-----KVLGIDTN-----FVTSKASDEGDPFELYL 957
 586 H-SFW-----SEYEVVLAWYAPPTVLALINHHKGFENVYAK-----GGLVEEY 630
 958 FKWTLLIPPTLLIYNLVGVVAGVSAINSQSGMPLFGKLFPAFWIVHLYPELKL 1017
 631 VDM-VISRPYIFLVLNLGVANGI-----WRIFYG----- 660
 1018 MGRONRPTIVVWSVLLASISFLMW 1044
 661 -----PPT-----EMLVVSMVWV 675
 RESULT 4
 BCSA_SALTY STANDARD; PRT; 874 AA.
 ID BCSA_SALTY
 AC Q93IN2;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).
 GN BCSA OR STM3619.
 OS (Salmonella typhimurium)
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella
 NC NCB1_TaxID-602;
 RN NC
 RX SEQUENCE FROM N.A.
 RP STRAIN-ATCC 14028;
 RX MEDLINE-21160181; PubMed-11260463;
 RA Zogaj X, Nimtz M., Rohde M., Bokranz W., Roemling U.;
 RT "The multicellular morphotypes of Salmonella typhimurium and
 RT Escherichia coli produce cellulose as the second component of the
 RT extracellular matrix";
 RL MOL. Microbiol. 39:1452-1463(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT72;
 RX MEDLINE-21927388; PubMed-11929533;

RA Solano C., Garcia B., Valle J., Beraaain C., Ghigo J.-M., Gamazo C.,
 RA Lasa I.,
 RT "Genetic analysis of *Salmonella enteritidis* biofilm formation:
 RT critical role for cellulose".
 RL Mol. Microbiol. 43:793-808(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE-21534948; PubMed-11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., All J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvoney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 LT2".
 RL Nature 413:852-856(2001).
 CC -1- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes
 CC uridine 5'-diphosphate glucose to cellulose, which is produced as
 CC an extracellular component for mechanical and chemical protection
 CC at the onset of the stationary phase. When the cells exhibit
 CC multicellular behavior (rod morphology). Co-expression of
 CC cellulose and thin aggregative fimbriae leads to a hydrophobic
 CC network with tightly packed cells embedded in a highly inert
 CC matrix.
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + ((1,4-beta-D-glucosyl))_(N) = UDP
 CC + ((1,4-beta-D-glucosyl))_(N+1).
 CC -1- COFACTOR: Magnesium (By similarity).
 CC -1- ENZYME REGULATION: Activated by bis-(3',-5') cyclic diguanylic acid
 CC (C-di-GMP) (By similarity).
 CC -1- PATHWAY: Bacterial cellulose biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Potential).
 CC -1- DOMAIN: There are two conserved domains in the globular part of
 CC the protein: the N-terminal domain (domain A) contains
 CC the conserved DXD motif and is possibly involved in catalysis and
 CC substrate binding. The C-terminal domain (domain B) contains the
 CC QXXRW motif and is present only in processive glycosyl
 CC transferases. It could be involved in the processivity function of
 CC the enzyme, possibly required for holding the growing glycan chain
 CC in the active site.
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ315770; CAC44015.1; -
 DR EMBL: AE008867; AAL22479.1; -
 DR EMBL: AJ315148; CAC86199.1; -
 DR StyGene: SG72727; bcsA.
 DR InterPro: IPR005150; Cellulose_synth.
 DR InterPro: IPR001173; Glycos_transf-2.
 DR Pfam: PF03552; Cellulose_synth; 1.
 DR Pfam: PF00535; Glycos_transf-2; 1.
 KW Cellulose biosynthesis; transferase; glycosyltransferase;
 KW Transmembrane; Inner membrane; Complete proteome.
 FT DOMAIN 271 364
 FT DOMAIN A CATALYTIC SUBDOMAIN A.
 FT TRANSMEM 30 50
 FT TRANSMEM 151 171
 FT TRANSMEM 173 193
 FT TRANSMEM 230 250
 FT TRANSMEM 525 545
 FT TRANSMEM 547 567
 FT TRANSMEM 592 612
 FT TRANSMEM 634 654
 FT TRANSMEM 668 688
 FT TRANSMEM 833 853
 FT TRANSMEM POTENTIAL.

FT ACT_SITE 313 313 POTENTIAL.
 FT ACT_SITE 457 457 POTENTIAL.
 FT SITE 360 360 SUBSTRATE BINDING (POTENTIAL).
 FT SITE 362 362 SUBSTRATE BINDING (POTENTIAL).
 SQ SEQUENCE 874 AA: 100043 MW: 409421B5806310A CRC64;
 Query Match 4.8%; Score 271; DB 1; Length 874;
 Best Local Similarity 18.8%; Pred. 1.2e-11;
 Matches 153; Conservative 90; Mismatches 202; Indels 370; Gaps 29;
 Oy 260 MYIMRLVILCLELHYRTNPV---PNAFALMVSVICELWALSMILDPKWPVNR 315
 Db 201 MLIVSLTVSCRYIMWRYSFLNMDDPVSIVCGILLFAETVAMIVLVGYEQVWPLNR 260
 Oy 316 EYIDRLALRYRBEPSQLAVDIFVSVVDLKEPPLVTANVTLSILAVDPVQVSCY 375
 Db 261 QP-----VLPKREMSQWPTDVFVPLYN---EDLVNKNVTIYASIGIDPKKLNW 309
 Oy 376 VPDGAAMLSFESLAESEFARKWVPCFKYSIEBRADWYPAKIDYLDKQVTSFVND 435
 Db 310 IIDDGGR---ESF---RHFAR----- 324
 Oy 436 RRAMKREVEEFKIRINALVSKALCPRESGWQDCTPWPGNNTGDHPGIVQLQNGGL 495
 Db 325 -----HVGVA----- 329
 Oy 456 DAGENDELRLVYVSREKRPQHHKAGAMNLVVSANVLTGPRILNDCDHYINNSKA 555
 Db 330 -----YIAFTT---HEHAKAGINNALNHA---KGEVVALFDCHVPTRSFL 370
 Oy 556 IREAMCFIMDPNIGQVQVDFPQRF---DGIDKN---DRYANRNVFEDINLGLDOI 608
 Db 371 QMTKMFLEK---KQLAMQRTPHHFSPDRERNLGRKRPNNGTGLFGLVQDGNDMW 426
 Oy 609 GCPVYVVGCGVFNPRALYGERPPYIKVHKKPSLSKLCGSSKKNKAKKESDKKSGNH 668
 Db 427 DATPFCGSAVIRR-----KP----- 442
 Oy 669 TDSYTPVFNLDIEGVEGAGFDDEKALLMSQMSLEKRGQSAVVASTLMENGVPESA 728
 Db 443 -----LDEI-----GGI----- 449
 Oy 729 TPENLKEAIVHISGVEDKSDGMELIGVSYEDILTGMKHAQRSIYCMPLKA 788
 Db 450 -----AVE-----TYEDAHITSLRLRGTSMV--RIQ 478
 Oy 789 FKSGAPINLSRLNOLKRNALGSVEILFSRHCPYGVNGRLKFLERFAYVTTIYPTS 848
 Db 479 AAGLATESLSAHIGIRIRARGMVOI-FRLDNPL--FGKGLKLAORLCYLNAMPHFISG 534
 Oy 849 IFLAKCT-----LANCLFTNQFIIPQI-----SNASIMFLSLSTFA 889
 Db 535 IPRLLFPLAPLAFELLHAYIYAPALMALFVIPHVHVASLTNSKIQGRHSFSEIYE 594
 Oy 890 TGILEMRNSGVGIDEMWRNEQFVIGVSAHLEAVFOGILKVLADITFTTTRASDED 949
 Db 595 TVLA-----WYIAPPI-----LVALINHKGFNFYDK----- 622
 Oy 950 GDPAEVLYEKWTLIPPTLLIVLVGVVAGVSAVNSGYOSMGPLFGLFPAFWIVH 1009
 Db 623 GGLVEKQYVDM-VISRPYFVLVLNLGLVAGV-----WRYVYQ----- 661
 Oy 1010 LYPFLKGLMGRNRPFTIVVWSVLLASIFSLNV 1044
 Db 662 -----ENETLTIV-----SLWV 675
 RESULT 5
 BCSA_XANAC STANDARD; PRT; 729 AA.
 ID BCSA_XANAC
 AC P58932;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)

FT	ACT SITE	337	337	POTENTIAL.
FT	SITE	240	240	SUBSTRATE BINDING (POTENTIAL).
FT	SITE	242	242	SUBSTRATE BINDING (POTENTIAL).
SO	SEQUENCE	729 AA:	80915 MW;	B9C08BB95E795B1 CRC64:
	Query Match	4.4%;	Score 252.5;	DB 1; Length 729;
	Best Local Similarity	19.6%;	Pred. No. 2e-10;	
	Matches 153;	Conservative 85;	Mismatches 212;	Indels 331; Gaps 28;
OY		260	NVIMRLVITLLEHYRTNP--VPNA--PALMLVSVCIEIWFALSMILDDGFKKPPNVR	315
Dd		81	MMGMSIAVSQRYIMRMQTOMGVGSADVFLIGLLGAELYAFAFIVLVGYFVLMPINR	140
OY		316	EYLDRLLRDRREGESQLAADVFESTVPLKEPPLVTANTVLSILAVYPVDKVCSCY	375
Dd		141	KP---VLPADQLRIMP---VDVEFITYN---EPISVTRTYLAASVIDMFAKIKIH	189
OY		376	VFDGAAMLSEFSIAETSEPAFKRVVPCKKYSIEBRADEWYPAKIDYLKOKVOTSFYKD	435
Dd		190	LIDDG-----	194
OY		436	RRAKKREEEKIRINALVSKALKCPREGGWMOGSTPWGNNTGPHGMIOVFLQGNGL	495
Dd		195	--RDERFRP-----CAEG-----	207
OY		496	DAEGNELPRLYVSVSEKRPFOHKKAGAMNALVKSVALTNGPEILNDCHYHNNSKA	555
Dd		208	-----INVYRTNNA-----HAKACINIMALK-----KSGDVAIFDCDH-IPTSF	249
OY		556	LREAN-CGLMPNIGKQCYQYQFORF-----DGIDKNDRYANRNTPEDLNLRGIDG	607
Dd		250	LQVAMGWLHDTKL---ALVQMHPYFSSPDPERFNIDTHKVPNEGELFFGLDDGDNDQ	305
OY		608	IQGPVVVGTCGVNFNTALYGYPEPIKYVKKRPSLLSKLGCSRRKNSKAKESDKKSGR	667
Dd		306	WMATFEFGSCAVIKRTA-----	322
OY		668	HTDSTVPVNLDDIEEGVEGAGFDDKALLMSQSLERKFGOSAFVASTLMENGCVPS	727
Dd		323	-----LEEYGV-----	329
OY		728	ATPENLKEAIHVISCGEYEDKSDMGMEIGWYGSYTEDLIGFKMHARGMRSTYCMPLP	787
Dd		330	-----AVE-----YTTEAHNRALKIQRGYKRAYL-AVP	357
OY		788	AFKGSAPINLSDRLNOVLRMALGSYEI-----LESRHCPITYGNGRLKEPFAVYNT	841
Dd		358	QAAGIATESISGHVQRIRMARMAQIARINDPILGR-----GLKTSORLCYINA	407
OY		842	TIYPTSIPILMYCTCLAVCFETNQFIIPQINSIASIWFLSILSIFATGI-ILEMRMSGV	900
Dd		408	MLHFPGYPRITLIAPLAYLFPGAHHVDAQSALMLAIALPHILOANTLNLSVOSRFNHL	467
OY		901	GIDEMWRNEQWVIGVSAHLEFAVEOGLIKLAGIDT-NFTVYSTKA-----SDEGDEAE	954
Dd		468	LMNEYETTAMNYI-----FRPLVALLNPKICKFMNVYGGVLARSYFDAQIAK	517
OY		955	LYLKRWITLLIPPTTLILVNLYGVAYG---SYAINSGYQS--WGPLEGLLFPAAVTVH	1005
Dd		518	PLYL-----LILLNVGMVAGVLRILIYVSGSEQDTIW-----ENIAMLTLYN	559
OY		1010	L 1010	
Dd		560	M 560	
	RESULT 6			
	ACS2_ACEXY			
AC	ACS2_ACEXY	STANDARD;	PRT,	1596 AA.
AC	059167:			
DT	15-JUN-2002	(Rel. 41, Created)		
DT	15-JUN-2002	(Rel. 41, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		

DE Cellulose synthase 2 (includes: Cellulose synthase catalytic subunit
DE [UDP-forming] (EC 2.4.1.12); Cyclic di-GMP binding domain (Cellulose
DE synthase 2 regulatory domain)).
GN ACSA11.
OS Acetobacter xylinus.
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconacetobacter.
OX NCBI_TaxID=28448;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23769;
RX MEDLINE=95394846; Pubmed=7665515;
RA Saxena I.M., Brown R.M. Jr.;
RT "Identification of a second cellulose synthase gene (acsa11) in
RT Acetobacter xylinum.";
RL J. Bacteriol. 177:5276-5283(1995).
CC -1- CATALYTIC ACTIVITY: UDP-glucose + ((1,4-beta-D-glucosyl)) (N) - UDP
CC + ((1,4-beta-D-glucosyl)) (N+1).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (potential).
CC -1- DOMAIN: There are two conserved domains in the globular part of
CC the catalytic subunit: the N-terminal domain (domain A) contains
CC the conserved DXD motif and is possibly involved in catalysis and
CC substrate binding. The C-terminal domain (domain B) contains the
CC OXXXR motif and is present only in processive glycosyl
CC transferases. It could be involved in the processivity function of
CC the enzyme, possibly required for holding the growing glycan chain
CC in the active site.
CC -1- MISCELLANEOUS: It is not essential for cellulose production in
CC this strain.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC GLYCOSYLTRANSFERASE FAMILY 2.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ACSB/BCSB
CC FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
CC or send an email to license@slb-slb.ch).
CC -----
DR EMBL, U15957; AAA85264.1; -
DR InterPro: IPR003919; Cellsynth_A.
DR InterPro: IPR003920; Cellsynth_B.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF03170; BcsB; 1.
DR Pfam: PF00535; Glycos_transf_2; 1.
DR PRINTS: PR01439; CELLSTHASEA.
DR PRINTS: PR01440; CELLSTHASEB.
DR Cellulose biosynthesis; Transferease; Glycosyltransferase;
KW Transmembrane; Inner membrane.
FT DOMAIN 1 749 CATALYTIC.
FT DOMAIN 750 1596 CYCLIC DI-GMP BINDING (POTENTIAL).
FT DOMAIN 145 238 CATALYTIC SUBDOMAIN A.
FT DOMAIN 315 375 CATALYTIC SUBDOMAIN B.
FT TRANSMEM 25 45 POTENTIAL.
FT TRANSMEM 106 126 POTENTIAL.
FT TRANSMEM 396 416 POTENTIAL.
FT TRANSMEM 421 441 POTENTIAL.
FT TRANSMEM 505 525 POTENTIAL.
FT TRANSMEM 544 564 POTENTIAL.
FT TRANSMEM 1553 1573 POTENTIAL.
FT ACT_SITE 187 187 POTENTIAL.
FT ACT_SITE 331 331 POTENTIAL.
FT SITE 234 234 SUBSTRATE BINDING (POTENTIAL).
FT SITE 236 236 SUBSTRATE BINDING (POTENTIAL).
SQ SEQUENCE 1596 AA; 175799 MW; D98A6F6259E1F3CE CRC64;

[illegible]

```

RN [1]
RP .SEQUENCE FROM N.A.
RC STRAIN-JCM 7664 / IFO 13693;
RX PubMed=10382968;
RA Umeda Y., Hirano A., Ishibashi M., Akiyama H., Onizuka T., Ikeuchi M.,
RT Inoue Y.;
RT Cloning of cellulose synthase genes from Acetobacter xylinum JCM
RT 7664: Implication of a novel set of cellulose synthase genes.*;
RL DNA Res. 6:109-115(1999).
CC -1- CATALYTIC ACTIVITY: UDP-glucose + ((1,4-beta-D-glucosyl))(N) = UDP
CC + ((1,4-beta-D-glucosyl))(N+1);
CC -1- PATHWAY: Bacterial cellulose biosynthesis.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -1- DOMAIN: There are two conserved domains in the globular part of
CC the catalytic subunit: the N-terminal domain (domain A) contains
CC the conserved DXD motif and is possibly involved in catalysis and
CC substrate binding. The C-terminal domain (domain B) contains the
CC QXXRW motif and is present only in processive glycosyl
CC transferases. It could be involved in the processivity function of
CC the enzyme, possibly required for holding the growing glycan chain
CC in the active site.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC GLYCOSYLTRANSFERASE FAMILY 2.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ACSE/BCSB
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB015803; BAA77593.1; -
DR InterPro: IPR003919; CellSynth_A.
DR InterPro: IPR001173; Glycosyl_transf_2.
DR Pfam: PF00535; Glycosyl_transf_2; 1.
DR PRINTS: PR01439; CELSINTHASE.
DR PRINTS: PR01440; CELSINTHASE.
KW Cellulose biosynthesis; Transferase; Glycosyltransferase;
KW Transmembrane; Inner membrane.
FT DOMAIN 1 731 CATALYTIC.
FT DOMAIN 732 1518 CYCLIC DI-GMP BINDING (BY SIMILARITY).
FT DOMAIN 144 237 CATALYTIC SUBDOMAIN A.
FT DOMAIN 314 374 CATALYTIC SUBDOMAIN B.
FT TRANSMEM 24 44 POTENTIAL.
FT TRANSMEM 71 91 POTENTIAL.
FT TRANSMEM 105 125 POTENTIAL.
FT TRANSMEM 404 424 POTENTIAL.
FT TRANSMEM 427 447 POTENTIAL.
FT TRANSMEM 465 485 POTENTIAL.
FT TRANSMEM 514 534 POTENTIAL.
FT TRANSMEM 543 563 POTENTIAL.
FT TRANSMEM 1481 1501 POTENTIAL.
FT ACT_SITE 186 186 POTENTIAL.
FT ACT_SITE 330 330 POTENTIAL.
FT SITE 233 233 SUBSTRATE BINDING (POTENTIAL).
FT SITE 235 235 SUBSTRATE BINDING (POTENTIAL).
SQ SEQUENCE 1518 AA; 166404 MW; 0EC99B35B6DE4543 CRC64;

```

Query Match 4.4%; Score 247.5; DB 1; Length 1518;
 Best Local Similarity 19.7%; Pred. No. 1.4e-09;
 Matches 152; Conservative 99; Mismatches 233; Indels 289; Gaps 30;

```

DB 133 -----EHDLPEDMAOWPSVDYFVPSYN---EELSLVRSYLGALDLMPADRLNV 181
OY 375 YFFDGAAMLSPESLAETSEFAKKVFPCKKYSIEPRADPEWFAAKIDYLNKDVQTSFYK 434
DB 182 YIIIDG-----
OY 435 DRRAMREYEERIRINALVSKALCPBEGWWDQTPWPGNNTGDPGMIOYFLQONGC 494
DB 188 -----RRKAFHDFAVEAGA-----GYIIR-----
OY 495 LBAEGNELRLVYVSEKRRPGQHHKAGAMALVYSAVLNTPPIILMDCHYINNSK 554
DB 207 -AENN-----HAKAONLNHALV-----FDSPPAVLFDH-VPTRG 241
OY 555 ALREACFLM-DENLKGVCYVOFPORFDGIDKNDRYANRNVYFDPINLRGLGIDGPY 613
DB 242 FLRRITGMMADPNL-----ALLQTPHIFAPDP-----FQRLNLAG-GMIVP-- 282
OY 614 VGTGCVFNRTALYGPPIKVKHKRPSLLSKLGGSRKKSKAKESDKKSRHTDSTV 673
DB 283 -PEGNMF-----YGL-----
OY 674 PVFNLDIDEGVBAQFDDERKALMSQSLERKFGQSAVFVASTLMENGVSPPATPVL 733
DB 292 -----VODGND-----FMDATFCCSCAI-----I 311
OY 734 LKEAIVHISGVDKSDGMETGMYSTEDILTFGKMHARGMRISYCPMKPAPKQSA 793
DB 312 RRAVVGIG-----GFATRYTEDAHNTLKNQRMGWAVLREPLAA--GLA 356
OY 794 PINLSRLNQLVRLWALGSVEILFSRHCPWYGVNGRLKLELRRPVAVNTTIYPTSLPLM 853
DB 357 TERLLIHIGQVRWARGMGIQIMRLDPMGLAG---LRWEOQRCYLSAMSHLFAIPRLT 412
OY 854 YCTLLAVCLFTNFFITPQSNIASIMEL-SLPLSTPATGILLEBMSGVDIDEMNRDEQF 912
DB 413 FLVSPPLAFLEGGNTITAAAPLAISYALPHIEFHSVTLSRTIGRW-----RYSF 462
OY 913 -VIGVSAHLFAVFOGILKVLGIDTNTFVTKASDEDDDFALYLFKWTTLIPPTTL 971
DB 463 SEIYERSLALFLRITITVLLQPHKKFNVT---DKGGLLAGRY-FDMDA-VYPAVILA 516
OY 972 IYNLVGVAVGSAVNSGYOSGGLFGLKFFAFWIVHLYPLKGL-MGRNR 1023
DB 517 GYLCAALLKGV-FGIWOFHDLALOSFTLMLVVISLIYLAISAVGREIR 568

```

RESULT 8
 BCAS_ACEXY STANDARD; PRT; 1518 AA.
 AC 09MX75;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative cellulose synthase 3 (includes: Cellulose synthase
 DE catalytic subunit (UDP-forming) (EC 2.4.1.12); Cyclic di-GMP binding
 DE domain (Cellulose synthase 3 regulatory subunit)).
 GN BCSABII-B.
 OS Acetobacter xylinus.
 OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
 OC Gluconacetobacter.
 OX NCBI_TaxID=28448;
 RN [1]
 RP .SEQUENCE FROM N.A.
 RC STRAIN-JCM 7664 / IFO 13693;
 RX PubMed=10382968;
 RA Umeda Y., Hirano A., Ishibashi M., Akiyama H., Onizuka T., Ikeuchi M.,
 RT Inoue Y.;
 RT Cloning of cellulose synthase genes from Acetobacter xylinum JCM
 RT 7664: Implication of a novel set of cellulose synthase genes.*;
 RL DNA Res. 6:109-115(1999).
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + ((1,4-beta-D-glucosyl))(N) = UDP
 CC + ((1,4-beta-D-glucosyl))(N+1);

CC the catalytic subunit: the N-terminal domain (domain A) contains
 CC the conserved DXD motif and is possibly involved in catalysis and
 CC substrate binding. The C-terminal domain (domain B) contains the
 CC QXXW motif and is present only in processive glycosyl
 CC transferases. It could be involved in the processivity function of
 CC the enzyme, possibly required for holding the growing glycan chain
 CC in the active site.
 CC - SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AB010645; BAA31463.1; -
 DR InterPro: IPR003919; CellSynth_A.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 DR PRINTS: PR01439; CELLSNTHASE.
 KM Cellulose biosynthesis: transferase; glycosyltransferase;
 KM Transmembrane, inner membrane.
 FT DOMAIN 147 242 CATALYTIC SUBDOMAIN A.
 FT TRANSMEM 319 379 CATALYTIC SUBDOMAIN B.
 FT TRANSMEM 27 47 POTENTIAL.
 FT TRANSMEM 49 69 POTENTIAL.
 FT TRANSMEM 106 126 POTENTIAL.
 FT TRANSMEM 167 187 POTENTIAL.
 FT TRANSMEM 409 429 POTENTIAL.
 FT TRANSMEM 432 452 POTENTIAL.
 FT TRANSMEM 470 490 POTENTIAL.
 FT TRANSMEM 517 537 POTENTIAL.
 FT TRANSMEM 551 571 POTENTIAL.
 FT ACT_SITE 189 189 POTENTIAL.
 FT ACT_SITE 335 335 POTENTIAL.
 FT SITE 238 238 SUBSTRATE BINDING (POTENTIAL).
 FT SITE 240 240 SUBSTRATE BINDING (POTENTIAL).
 SQ SEQUENCE 756 AA; 84561 MW; 6954F39A25E73B0A CRC64;
 Query Match 4.28; Score 236.5; DR 1; Length 756;
 Best Local Similarity 19.38; Pred. No. 3; Le-09;
 Matches 166; Conservative 98; Mismatches 263; Indels 331; Gaps 34;
 204 EKNTGCVSTQASERGVGVIDASTD---ILADEALLNDEAR-----OPL 244
 11 ESRIGISIKKILSLKRASTYVGLGICALIATTVTLNNEQLVAACVVFVVGKRG 70
 245 SRKVSIPSSRINRYRMVIMLRLVILCLFLHYRTNPVNAFALML-----SVSICEIMF 298
 71 SRRTOI-----FLEVLISALVSLRYLWRLTETLD-FNWIGIGILVILMELVA 119
 299 ALSWILDOPKRPVPRVRETYLDRLALRYREGSPSOLAANDIVSVDPKPEPLVTANT 358
 120 LKMLFSYOTQIQLRRAP---LPLPDVNDWMP---VDLPIFYD---EQLSIVRLT 168
 359 VLSILAVDPVVKSCYVFDGGAAMLSESLAETSEFARKWVFCCKYSIEPRAPMYFA 418
 169 VIGALGIDWPPDKVNYIILDG----- 190
 419 AKIDYLDKVVQVSFKDRRAKREVEFKIRINALVSKLCPREGWVMDGTTPMGNTT 478
 191 -----VRPEFQF-----AKDC----- 202
 479 GDHPGMIOYFLONGGLDAGNELRLVYVSRKRGFQHHKKAAGAMALVRSVAVLTNG 538
 203 ---GALYI---GAVDWD---SAHAKAGMLNAIK---RTSG 231
 539 PFLNADCDHYIINNSALREAMFLMDPNLKGQVYQFPPORDGIDKNDRYANRNTVEF 598
 232 DYILLIDCDH-IPTRAFLOIAMGMV---ADRKIALMQIPRHYSFDP----- 275

QY 599 DINKRLDGIQGFVYVGTGCVENRTALVYEPIKVKHKKPSILSKLGGSRKKSKAKK 658
 Db 276 -----FQRLAVGVYRTP----- 287
 QY 659 ESDKKSGRHTDSFVFNLDIEEGVEGAFDEKALLMSQMSLEKRFQSGAVFVASTL 718
 Db 288 -----PEGNL---FYGVIODGND-----FMDATFFGSCA 314
 QY 719 MNGVPPSATPENILKEAIVHISCGEDKSDWMEIGVYGVTEIDILTGFMARGMR 778
 Db 315 I-----LRRALIESIG-----GRAVEVETDATHALMQRGMS 348
 QY 779 STYCMKPLPAFKSGAPINISDRNLQVLRNALSVELLSRHPITWGYGR-LKFLERRA 837
 Db 349 TAYL--RIVASGLATERLTTHGQHMRAWMIOI-FRYVNDML---GRGKIKQQRIC 401
 QY 838 YVNTTIPITSTPLMWCYLLNCLTQNFPIQISNIMSIMWLSFLSIFANGILEMR 897
 Db 402 YLSAMTSFPPALPRVIFELASPLAFLPAGQNTIIPAALAVAAVALPHEHSIAL----- 454
 QY 898 SGVIDEMWNRNEQFW-VIGGVSALFAVFGILKVLGIDTNETVYSKA---SDEGDPA 953
 Db 455 -AAKVKKMR-YSPFSEVETTMALFLVAVTITVTLFPKSKFNTVEKGGVLEEFEDIG 512
 QY 954 ELY-----LFWTTLIPPTILLI-VNLGVAVGVYVAINSGVSGPLPKLPAFV 1006
 Db 513 ATVPNIFATIMGGILIGFELIVFNPOLDVYARNAYLNLCA-----WA 557
 QY 1007 IVHLVPLKGL-MGRQNR 1023
 Db 558 LISLILFAALAVGRETK 575
 RESULT 10
 ID BCAL_ACEXY STANDARD; PRT; 754 AA.
 AC P19449;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).
 GN BCSA.
 OS Acetobacter xylinus.
 OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
 OC Gluconacetobacter.
 OX NCBI_TaxID=28448;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-42 AND 196-206.
 RC STRAIN-1306-3;
 RX MEDLINE-91045951; Pubmed-2146681;
 RA Wong H.C., Fear A.L., Calhoun R.D., Eichinger G.H., Mayer R.A.,
 RA Amikam D., Benizman M., Gelfand D.H., Meade J.H., Emerick A.W.,
 RA Bruner R., Ben-Bassat A., Tel R.;
 RT "Genetic organization of the cellulose synthase operon in Acetobacter
 RT xylinum.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8130-8134(1990).
 RN [2]
 RP ENZYME REGULATION.
 RC STRAIN-1306-3;
 RX Pubmed-11297407;
 RA Chang A.L., Tuckerman J.R., Gonzalez G., Mayer R., Weinhouse H.,
 RA Volman G., Amikam D., Benizman M., Gilles-Gonzalez M.-A.;
 RT "Phosphodiesterase A1, a regulator of cellulose synthesis in
 RT Acetobacter xylinum, is a heme-based sensor.";
 RL Biochemistry 40:3420-3426(2001).
 CC - FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes
 CC uridine 5'-diphosphate glucose to cellulose in a processive way.
 CC The thick cellulosic mats generated by this enzyme probably
 CC provide a specialized protective environment to the bacterium.
 CC - CATALYTIC ACTIVITY: UDP-glucose + [(1,4)-beta-D-glucosyl](N) - UDP
 CC + [(1,4)-beta-D-glucosyl](N+1).
 CC - COFACTOR: Magnesium.
 CC - ENZYME REGULATION: Activated by bis-(3'-5') cyclic diuanylic acid


```

CC      (c-di-GMP) .
CC      -1- PATHWAY: Bacterial cellulose biosynthesis.
CC      -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC      (Potential) .
CC      -1- DOMAIN: There are two conserved domains in the globular part of
CC      the catalytic subunit: the N-terminal domain (domain A) contains
CC      the conserved DXD motif and is possibly involved in catalysis and
CC      substrate binding. The C-terminal domain (domain B) contains the
CC      QXXRW motif and is present only in processive glycosyl
CC      transferases. It could be involved in the processivity function of
CC      the enzyme, possibly required for holding the growing glycan chain
CC      in the active site.
CC      -1- INDUCTION: Cellulose is produced at a linear rate with respect to
CC      cell growth when O(2) is present.
CC      -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed, usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC
CC      -----
CC      DR      EMBL: M37202; AAA21884.1; .
CC      DR      PIR: A43735; A43735.
CC      DR      InterPro: IPR003919; CellulynthA.
CC      DR      InterPro: IPR001173; Glycos_transf_2.
CC      DR      Pfam: PF00535; Glycos_transf_2; 1.
CC      DR      PRINTS: PR01439; CELLUSMTHASEA.
CC      KW      Cellulose biosynthesis; transferase; glycosyltransferase;
CC      KW      Transmembrane; Inner membrane.
CC      FT      DOMAIN 147 240
CC      FT      TRANSNM 317 377
CC      FT      TRANSNM 26 46
CC      FT      TRANSNM 47 67
CC      FT      TRANSNM 108 128
CC      FT      TRANSNM 167 187
CC      FT      TRANSNM 407 427
CC      FT      TRANSNM 430 450
CC      FT      TRANSNM 468 488
CC      FT      TRANSNM 516 536
CC      FT      TRANSNM 549 569
CC      FT      ACT_SITE 189 189
CC      FT      ACT_SITE 333 333
CC      FT      SITE 236 236
CC      FT      SITE 238 238
CC      SO      SEQUENCE 754 AA; 84442 MW; 805FCFE82E2C066 CnC64;
CC
CC      Query Match 4.0%; Score 228.5; DB 1; Length 754;
CC      Best Local Similarity 19.3%; Pred. No. 1.2e-08;
CC      Matches 154; Conservative 84; Mismatches 229; Indels 329; Gaps 30;
CC
CC      Oy      263 MLRLVILLCFLHYVITNPVPAFAFLML-----VSVCIEIFALSMILDDPPKPEPNRE 316
CC      Db      80 VLSLVLSRLYTWLRTETLD--EPFWIOGLGLVTLLMAELYALMLFLSYFOTIOPFLRA 137
CC      Oy      317 TYLDRLALRYDEEPEPSOLAIVDFVSTDLKEPPVLTANTVYSLTAVDPYDKVSCYV 376
CC      Db      138 P-----LPLPDVNDWMT-----VDLEFIPTD---EQLSIVRLTIVLGALGIDMPPKVNYYI 186
CC      Oy      377 FDDCAAMLSESLAETSEFAKKWVPCKKKSIEPRADPEWFAAKIDYLUKDKVQTSFYKDR 436
CC      Db      187 LDDG-----
CC      Oy      437 RAKREYEERFIRINALVSKALCPDEEGWVWQDGPMPGNGNTGDHPGMIOYFLGONGSLD 496
CC      Db      191 --VAPPEQFAKDCGAL-----
CC      Oy      497 AEGNELPLRYVVSREKRPFOHKKAGAMNALVRVSAVLITNGPFLNLDCDHYINNSKAL 556
CC      Db      206 -----YIGAVD-----SSNAKKAGLNHAIR-----RTSGDYITLLDCDH-IPTRAF 246

```

Oy	557	REAKGFLMDPNLGRKVCVCFVFPQRDRDGDKNRBYANRNTVPFDINLRGLDGIOPVYVG	616
Db	247	QIANGMWN---ADRRIALMLQTPHNHFSPD-----	273
Oy	617	GCVENRTALYCEPPKIVKKHKRPSLSLKCGSRKKNSKAKKESDKKSGRHTDSVPYF	676
Db	274	---FQRNLAVGRTP-----	288
Oy	677	NLDIDIEGVEGAFDEKALMLMSQSLERFGQSAVFAVASTLMENGVPPSATPENILKE	736
Db	289	NL---FYGVIDQGN-----FMDATFFGSCAI-----LRRE	317
Oy	737	AIHYISGCEYEDKSMGMELIGVYSTDIEDLTGFKHAGKMSIYCPKLPFAKSGAPIN	796
Db	318	AIESG-----GFAVEVYTEDAHNLQRGRGSTAYL--RIPVASGLATER	362
Oy	797	LSDRNLQVLRNALGSEVLEIFSRCPHWGYNRKLRFLEBPAAVNTIYPTISPLMYG	856
Db	363	LTHIGQMRNARGTIQI-FRDNFM---LGGGLKQGLCYLSANTSEFFAIPRIYFLA	418
Oy	857	LLAVCLTFNOETIIPDISNIASIMFLSLFLEISFATGILEKRWSS---GVGIDEMWRNQFW-	912
Db	419	SPLAFLEFGQNTIASP-----LAVLAIPIPHNFHSIATAKVNKKMR-VSPMS	466
Oy	913	VIGCVSALHFAVFGCIGLKVLAGIDINFTVTSKA---SDSDGPAELYLEKWTLLIPPT	969
Db	467	EYETTMALFLVRVYIITLFFPSKKEFNTERKGVLDEEDEFDGAIV-----PNI	516
Oy	970	LLIYVLGVVAGVSAINSQVSGWQPLFG--KLFPFWYIYHLPLKGLMGQRNPTI	1022
Db	517	IF-----AGIMTLGLIGLFELTFHF-----NOLAGIAKRAYLL	550
Oy	1028	VVWSS-----VLLASI	1038
Db	551	NCIMAMISLIILLAAI	566
RESULT 11			
BCSA_PSEFL			
ID	BCSA_PSEFL	STANDARD:	PRT: 739 AA.
AC	p58931:		
DT	15-JUN-2002 (Rel. 41, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).		
GN	BCSA OR MSSB.		
OS	Pseudomonas fluorescens.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;		
OC	Pseudomonas.		
OX	NCBI_TaxID=294;		
FN	SEQUENCE FROM N.A.		
RP	STRAIN-SBM25:		
RX	MEDLINE=22013850; PubMed=12019221;		
RA	Spiers A.J., Kahn S.G., Bohannon J., Travisano M., Ratney P.B.:		
RT	"Adaptive divergence in experimental populations of Pseudomonas		
RT	fluorescens. I. Genetic and phenotypic bases of wrinkly spreader		
RT	fitness."		
RL	Genetics 161:33-46(2002).		
CC	-1- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes		
CC	uridine 5'-diphosphate glucose to cellulose, which is produced as		
CC	an extracellular component responsible for the structural		
CC	integrity and rigidity of self-supporting mats characteristic of		
CC	the "wrinkly spreader" phenotype.		
CC	-1- CATALYTIC ACTIVITY: UDP-glucose + ((1,4-beta-D-glucosyl))(N) - UDP		
CC	+ ((1,4-beta-D-glucosyl))(N+1).		
CC	-1- COFACTOR: Magnesium (By similarity).		
CC	-1- ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid		
CC	(c-di-GMP) (By similarity).		
CC	-1- PATHWAY: Bacterial cellulose biosynthesis.		
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane		
CC	(Potential).		
CC	-1- DOMAIN: There are two conserved domains in the globular part of		

CC the protein: the N-terminal domain (domain A) contains
 CC the conserved DXD motif and is possibly involved in catalysis and
 CC substrate binding. The C-terminal domain (domain B) contains the
 CC OXNRM motif and is present only in processive glycosyl
 CC transferases. It could be involved in the processivity function of
 CC the enzyme, possibly required for holding the growing glycan chain
 CC in the active site.
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: AY074776; AAL71842.1; -
 KW Cellulose biosynthesis; Transferase; Glycosyltransferase;
 KM Transmembrane; Inner membrane.
 FT DOMAIN 157 250 CATALYTIC SUBDOMAIN A.
 FT TRANSMEM 327 387 CATALYTIC SUBDOMAIN B.
 FT TRANSMEM 36 55 POTENTIAL.
 FT TRANSMEM 83 76 POTENTIAL.
 FT TRANSMEM 101 138 POTENTIAL.
 FT TRANSMEM 116 138 POTENTIAL.
 FT TRANSMEM 417 436 POTENTIAL.
 FT TRANSMEM 440 462 POTENTIAL.
 FT TRANSMEM 524 546 POTENTIAL.
 FT TRANSMEM 551 573 POTENTIAL.
 FT TRANSMEM 199 199 POTENTIAL.
 FT ACT_SITE 343 343 POTENTIAL.
 FT SITE 246 246 SUBSTRATE BINDING (POTENTIAL).
 FT SITE 248 248 SUBSTRATE BINDING (POTENTIAL).
 SQ SEQUENCE 739 AA; 82165 MW; 28962EA3854B2BB CRC64;
 Query Match 4.0%; Score 228; DB 1; Length 739;
 Best Local Similarity 18.8%; Pred. No. 1.3e-08;
 Matches 149; Conservative 87; Mismatches 213; Indels 342; Gaps 33;
 257 PYRWIMLRVLTIC-----FLHYRTNPVNPAPALML-----VSICEIMFAL----- 300
 DB 80 PERLAI-LALVLSLVSLSLRIFMRLITSL--GFETWDMFGYGLVAEFYALVLTIG 136
 QY 301 ---SWILDQFKMPVNRRETYLRLALRYDREGPSQLAAYDIFSVYDPLKEPPLVTA 356
 DB 137 YVQTAMPLRRTPEVML-----KTEPEEMPTVDVFIPTYN--EALSIYK 176
 QY 357 NTVLSILANDYVDKVCYVEDGAAMLSFEISLAIESFARKWVPCKKYSIEPRAPENY 416
 DB 177 LTFEAAQAMDWPKDKLRVHVLDDG-----RRDDEFE--FCRKVGYN----- 215
 QY 417 FAAKIDYLDKVOYTFVKDRBRAMKREVEEFKIRINALVSKALCPREGVMODGTPWPGN 476
 DB 216 ----- 215
 QY 477 NTGDHPGMIGVFLGONGGLDAEGNELPRLVYVSREKRPGEFOHKKAGANALVRYSAYLT 536
 DB 216 -----YIRRDV-----NPHAKAGNLNEALKV-----T 237
 QY 537 NCPFLINDCDHYINNSKALREAMGFLMDPNIGKOVQYQFQRF--DGIKRN-DRY-- 590
 DB 238 DGEYIALDPADHVPTRSEFLQVSLGFWLKPRL-----AMQTPHFESDPPEKRLDTPRA 293
 QY 591 -ANRNTVFEDINLRGLDGIQGVYVGTGVFNRTALYGEPRPIKVKHKRPSLSKLCGS 649
 DB 294 VNEBEELFGVQDNDLMAATFFGSCAVYR-----EP----- 328
 QY 650 RKKNSKAKKESDKKSGRRHTDSTVPFNLDIEEGVEGAFDEKALIMSQMSLEKRFQ 709
 DB 329 ----- 328
 QY 710 SAVFVASTLMENGVPSPATPENLKAHIVISCGYEDKSDMGMEIGWYGSVTEDILT 769

DB 329 -----LLEIGV-----AVE-----YTDEAHTA 347
 QY 770 FPMHANGMSIYCMPLPAFKSAPINLSDDLNOVLRMAIGSEVILFSRCHPLWGYNR 829
 DB 348 LKLNRIQYNTALV--AIPQAGLATESLSRHRNORIRMRGAQIFRTDPLLGK---- 401
 QY 830 LKLERFAYVNTIYPTISIPILMYCTL-LAVCLEPTNPIIPIQISNIASINWELSLT-SI 887
 DB 402 LKMGORICVANNMQHFFGLPLVFLTAPLALVING-----AEIFASALMAYAVLPHL 456
 QY 888 FATGILEMRSGVGIDEMRNQF-----WVIGVSAHLFAVFOGLKLKLAGIDNFTYT 942
 DB 457 VHSSTLNSRIQGRFRHSFV-NEVEYETVLAWYI-----LPVYVALVNPKAG--GFNVT 506
 QY 943 SKASDEGDFAELIEFKWTLTLLIPTTLIVLVGVAGVSAVINSYGSMGLGKLPF 1002
 DB 507 DKGIIDKQF-----FDW-KLARPYLVLAIVNLIGFGIHLI-----WGDASTAVTV 554
 QY 1003 AF---WVIVHL 1010
 DB 555 AINLFWTLVNL 565
 RESULT 12
 ID BCAS3_ACEXY STANDARD; PRT; 745 AA.
 AC 09WX61;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cellulose synthase 1 catalytic subunit [UDP-forming] (EC 2.4.1.12).
 GN BCSAI.
 OS Acetobacter xylinus.
 OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
 OC Gluconacetobacter.
 OX NCBI_TaxID=28448;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JCM 7664 / IFO 13693;
 RX PubMed=10382968;
 RA Umeda Y., Hirano A., Ishibashi M., Akiyama H., Onizuka T., Ikenuchi M.,
 RA Inoue Y.;
 RA "Cloning of cellulose synthase genes from Acetobacter xylinum JCM
 RT 7664: implication of a novel set of cellulose synthase genes";
 RL DNA Res. 6:109-115(1999).
 CC -1- FUNCTION: Catalytic subunit of cellulose synthase. The thick cellulosic
 CC matrix generated by this enzyme probably provide a specialized
 CC protective environment to the bacterium (By similarity).
 CC + ((1,4)-beta-D-glucosyl)(N+1).
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + ((1,4)-beta-D-glucosyl)
 CC + ((1,4)-beta-D-glucosyl)(N+1).
 CC -1- COFACTOR: Magnesium (By similarity).
 CC -1- ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid
 CC (c-di-GMP) (By similarity).
 CC -1- PATHWAY: Bacterial cellulose biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Potential).
 CC -1- DOMAIN: There are two conserved domains in the globular part of
 CC the catalytic subunit: the N-terminal domain (domain A) contains
 CC the conserved DXD motif and is possibly involved in catalysis and
 CC substrate binding. The C-terminal domain (domain B) contains the
 CC OXNRM motif and is present only in processive glycosyl
 CC transferases. It could be involved in the processivity function of
 CC the enzyme, possibly required for holding the growing glycan chain
 CC in the active site.
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AB015802; BAA77565.1;
DR InterPro; IPR003919; CellSynth_A.
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf.2; 1.
DR PRINTS; PR01439; CELLSTNTHASEA.
KW Cellulose biosynthesis; Transferase; Glycosyltransferase;
FT Transmembrane, inner membrane.
FT DOMAIN 147 240
FT DOMAIN 317 377
FT TRANSMEM 29 49
FT TRANSMEM 106 126
FT TRANSMEM 153 173
FT TRANSMEM 407 427
FT TRANSMEM 430 450
FT TRANSMEM 468 488
FT TRANSMEM 515 535
FT TRANSMEM 547 567
FT TRANSMEM 649 669
FT ACT_SITE 189 189
FT ACT_SITE 333 333
FT SITE 236 236
FT SITE 238 238
SQ SEQUENCE 745 AA; 83518 MM; 57EA0457A226F815 CRC64;

Query Match 3.7%; Score 208.5; DB 1; Length 745;
Best Local Similarity 18.9%; Pred. No. 3.3e-07;
Matches 149; Conservative 82; Mismatches 235; Indels 321; Gaps .31;

OY 263 MLRLVILCLFLHYRTNPVNAFLM-----LVSVICEIMALSMILDQPKMPPVRE 316
DB 80 VLSALVSLRYLFWRLFTLTD--FDTWGCGILGVTLLLELVALYMLFSLYFOTISPLHRA 137
OY 317 TYLDRLALRYDREGESQOLAVDIEFSTVDPLKEPPLTANTLVLAVDYVDKVCYV 376
DB 138 P-----LPL-----PANPDEMPYVDIFITYD--ELSLYRLVLGALGIDMPDPKVNVI 186
OY 377 FDDGAAMLSFESLAESEPAKRWPKCKKYSTIEPAPMEYFAKIDYLKDKVQTSFVKDR 436
DB 187 LDDG-----RREFAR-----FA----- 199
OY 437 RAMKREYEFKIRIALVSKALKCEBGMVMDGTPWPGNNNGDHPGMIOVFLGONGID 496
DB 200 -----E 200
OY 497 AEGNELPRLVYVSRKRPQHHKAGAMNALVRSAYLTNGPFIILNDCDHYNNSKAL 556
DB 201 ACGAR-----YIARDNA-----HAKAGNLNTAIK-----HTTGDHILILDCDH-ITPRAFL 246
OY 557 REAMCELPDNLGKOVYVQRPQRFQIDKNDRYANRNTVPFDINLRGLDIGIQGVYVGT 616
DB 247 QISMGMWVSDS-----NIALLOTPHNHFSPD----- 273
OY 617 GCVENRRLVGYEPIRYKHKKPSLLSKLGGSRKKSKAKKESDKSGRTDSTVPVF 676
DB 274 ---FORNLAVGTRP-----PEG 288
OY 677 NLDDIEGVEGAFDEKALMSOMSLERFGOSAVFVASTLMENGVPSPATPENLKE 736
DB 289 NL---FYGVLDGND-----FWDAFFFGSCAI-----LRK 317
OY 737 AIHWISCGYEDKSDGMEIGWYGVTEDIITGFEMHARGWSIYCMPLPAFKGSAPIN 796
DB 318 AIEEIG-----GFATETVETDAHTALRMQRGWTAVL--RIPLASGLATER 362
OY 797 LSDRLNOVLRNALSGVELLSRHCIPWYNGYRNRLKFLERFAVNTIYPIISPLMCT 856
DB 363 LITHIGQRKRRARGTQI--FRVDNPM---LGGSLKLGQCLCLASMTSFEFFAIPVITFLA 418
OY 857 LLAVCLFTNOFTIPIQISNIASIWFLSLFLSFATGILERMWGSVGIDEWMRNEQFW--VI 914

```

```

DB 419 SPLAEFLFSQNTIITAA SPLAVGYAIPHMFHSIAT-----AAKYNKGR-YSEWSEVY 469
OY 915 GGVSAHLRAVFGGILAKVLADIDTNTFYNSKA-----SDEGDGFALYLFKRTTLLIPTTL 971
DB 470 ETVMA-LFLVRYVTITMLFSPKGFNVEKGVLEKEEEDLATY-----PNIIPA 519
OY 972 IYNLVGVVAGV-----SYAINSGQSMGFLGKLFAFWYVHLYPLKGL 1017
DB 520 IIMALGLGLYALLFQHLDIISERAYALN-----CIWVSILITIMAVI 564
OY 1018 -MGRNR 1023
DB 565 SVGREK 571

RESULT 13
ID ACSE1 ACSEXY STANDARD; PRT; 1550 AA.
AC P21877; P37717;
DT 01-MAY-1991 (Rel. 18, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulose synthase 1 [includes: Cellulose synthase catalytic domain
DE [UDP-forming] (EC 2.4.1.12); Cyclic di-GMP binding domain (Cellulose
DE synthase 1 regulatory domain)].
CN ACSAB OR ACSA OR ACSB.
OS Acetobacter xylinus.
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconobacter.
OX NCBI_Taxid=28448;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 6-20.
RC STRAIN=ATCC 53582;
RA MEDLINE=91346705; PubMed=2151718;
RX Saxena I.M., Lin F.C., Brown R.M. Jr.;
RT "Cloning and sequencing of the cellulose synthase catalytic subunit
RT gene of Acetobacter xylinum.";
RL plant Mol. Biol. 15:673-683(1990).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 768-781.
RC STRAIN=ATCC 53582;
RA MEDLINE=91322509; PubMed=1830823;
RX Saxena I.M., Lin F.C., Brown R.M. Jr.;
RT "Identification of a new gene in an operon for cellulose biosynthesis
RT in Acetobacter xylinum.";
RL plant Mol. Biol. 16:947-954(1991).
RN [3]
RP REVISIONS.
RC STRAIN=ATCC 53582;
RX PubMed=8083166;
RA Saxena I.M., Kudlicka K., Okuda K., Brown R.M. Jr.;
RT "Characterization of genes in the cellulose-synthesizing operon (acs
RT operon) of Acetobacter xylinum: implications for cellulose
RT crystallization.";
RL J. Bacteriol. 176:5735-5752(1994).
RN [4]
RP SEQUENCE OF 1-8 FROM N.A.
RC STRAIN=ATCC 23769;
RX MEDLINE=94131945; PubMed=8300521;
RA Standaal R., Iversen T.-G., Coucheron D.H., Fjaervik E., Blatny J.M.,
RA Valla S.;
RT "A new gene required for cellulose production and a gene encoding
RT cellulolytic activity in Acetobacter xylinum are colocalized with the
RT bcs operon.";
RL J. Bacteriol. 176:665-672(1994).
RN [5]
RP FUNCTION.
RC STRAIN=ATCC 53582;
RX PubMed=2138620;
RA Lin F.C., Brown R.M. Jr., Drake R.R. Jr., Haley B.E.;
RT "Identification of the uridine 5'-diphosphoglucose (UDP-Glc) binding
RT subunit of cellulose synthase in Acetobacter xylinum using the
RT photoaffinity probe 5-azido-UDP-Glc.";

```


QY 958 FKWTTLLIPTLLIVNLGVVAGVSYAINSGYSGWPLFGKLPFAFWIVLHLPKGL 1017
 Db 510 -----GAVYPIILGLIMFGGLARGV-YELSGHLDQIAERAYLLNSAMMLSTIIILAAI 564
 QY 1018 -MGRO-----NRPPIVVV 1030
 Db 565 AVGRETOOKRNSHRIPIPV 585

RESULT 14
 COX1_PARLI STANDARD; PRT; 517 AA.
 ID COX1_PARLI STANDARD; PRT; 517 AA.
 AC P12700;
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
 GN COI.
 OS Paracentrotus lividus (Common sea urchin).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidae; Euechinoidea; Echinacea; Echinoida; Echinidae;
 OC Paracentrotus.
 OX NCBI_TaxID=7656;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-69291831; PubMed-2544576;
 RA Cantatore P., Roberti M., Rainaldi G., Gadaleta M.N., Saccone C.;
 RT "The complete nucleotide sequence, gene organization, and genetic
 code of the mitochondrial genome of Paracentrotus lividus.";
 RL J. Biol. Chem. 264:10965-10975(1989).
 RN [2]
 RP SEQUENCE OF 469-517 FROM N.A.
 RC TISSUE-Bgg;
 RX MEDLINE-67248108; PubMed-3596250;
 RA Cantatore P., Roberti M., Morisco P., Rainaldi G., Gadaleta M.N.,
 RA Saccone C.;
 RT "A novel gene order in the Paracentrotus lividus mitochondrial
 genome.";
 RL Gene 53:41-54(1987).
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
 CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 AND COPPER B.
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
 c + 2 H(2)O.
 CC -1- PATHWAY: Respiratory chain; terminal step.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J04815; AAA68135.1; -
 DR EMBL: M16519; AAA31989.2; -
 DR PIR: C34284; C34284.
 DR HSSP: P00396; 20CC.
 DR InterPro: IPR000883; COX1.
 DR Pfam: PF00115; COX1.1.
 DR PRINTS: PR01165; CYCOXIDASE1.
 DR PROSITE: PS00077; COX1.1.
 DR Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
 KW Respiratory chain; Inner membrane.

FT METAL 61 61 IRON (HEME A) (PROBABLE).
 FT METAL 240 240 COPPER B (PROBABLE).
 FT METAL 244 244 COPPER B (PROBABLE).
 FT METAL 290 290 COPPER B (PROBABLE).
 FT METAL 291 291 COPPER B (PROBABLE).
 FT METAL 376 376 IRON (HEME A3) (PROBABLE).
 FT METAL 378 378 IRON (HEME A) (PROBABLE).
 SQ SEQUENCE 517 AA; 57217 MW; B42A1P1D5789A856 CnC64;

Query Match 2.1%; Score 119; DB 1; Length 517;
 Best Local Similarity 20.3%; Pred. No. 0.63;
 Matches 87; Conservative 56; Mismatches 139; Indels 146; Gaps 21.

QY 724 VPPSATPENLKEAIVHISGYEDKSDGMELG-IVGSYEDI-----LNGFKHA 774
 Db 105 IPPSP-----ILLASAGVES---GACTGWTITPPLSSNIAHAGSVDLAIFSJHL 152
 QY 775 RGMRSIYC-----MPKLPAGSAPINLSD 799
 Db 153 AGASSILASINFTTIIMRTPGMSFDRLPFWMSVFTAPFLLLSLPVLGATIMLLTD 212
 QY 800 R-----LNQVLRNALGSVEL-----FSRHCPITWGYNGRLFLERFA 837
 Db 213 RNINTTFEDPAGGDDPIFHQLFWFEGHPEVYIILPFGMISHVIAHYSK---REDFG 269
 QY 838 YVNTIYPTISPIPLMCTLLAVCLFTNQFIIPQISNASTWFLSTFIATGILEMRW 897
 Db 270 YLG-MVYAMIAIGVGF-LVMAHMF---GMDVTRAYTAAMITIAV-PTGIKPSW 323
 QY 898 SGV--GIDEMWRNEQFVIGCVSAHLFAVFOGILKVLGIDTNTFTVTSKASDEGDFAEL 955
 Db 324 MATIAGSNLQWETPLMALGIFVPLETGLGLVIANSD-----GKL-PPAF 1004
 QY 956 YLFKWTLLIPTLLIVNLGVVA-GSVAINSGYSGWPL-----GKL-PPAF 1004
 Db 365 -----VLHDITYVVAHFHYVLSMGAVFAIFAGFTHMPFCGYNLHPLMGKHFPM 417
 QY 1005 WVIHLYPF-----LKLGMRONRTPPIVVMSV-----LLASIF--SLWVRID 1047
 Db 418 FVGYNLTFPPQHFGLAGMPRYSDDPATLMTNVSSIGSTISLVANLFFFLIW---E 474

QY 1048 PFSRWVG 1055
 Db 475 AFASQREG 482

RESULT 15
 YKTA_CAEEL STANDARD; PRT; 590 AA.
 ID YKTA_CAEEL STANDARD; PRT; 590 AA.
 AC P34322;
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein C07A9.11 in chromosome III.
 GN C07A9.11.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 CC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnscough R., Anderson K., Baynes C., Beirs M.,
 RA Bonfield J., Burton J., Connell M., Cosey T., Cooper J., Coulson A.,
 RA Crixton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawley T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkhan R.,
 RA Sims M., Smaildon N., Smith A., Smith M., Sonnenhammer E., Staden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,


```

Db 61 SCPOCKTRKRLKGSALIPGDDKDEDLAEGTVEFVYPOKEKISERMILMHLTRGGEEM 120
QY 121 GEPQYDKESVHNLPLRLTSRODTSGEFSAASPERLSVSTIAGKRLPYSSDNOGPNRR 180
Db 121 GEPQYDKESVHNLPLRLTSRODTSGEFSAASPERLSVSTIAGKRLPYSSDNOGPNRR 180
QY 181 IYDPVGLGNVAMKERVYDGMKMKOEKNTGVSVOAASERGVDIDASTDILADEALLNDEA 240
Db 181 IYDPVGLGNVAMKERVYDGMKMKOEKNTGVSVOAASERGVDIDASTDILADEALLNDEA 240
QY 241 ROPLSKRVSSIPSSRINPYRMVIMLRVYICLFHYRITNPVPAFALMLVSYCEIEMFAL 300
Db 241 ROPLSKRVSSIPSSRINPYRMVIMLRVYICLFHYRITNPVPAFALMLVSYCEIEMFAL 300
QY 301 SWILDOFPKMFVFNRETYIDRLALRYDRGEGESQOLAVIDIFSTVDPLEKPLVTANTYL 360
Db 301 SWILDOFPKMFVFNRETYIDRLALRYDRGEGESQOLAVIDIFSTVDPLEKPLVTANTYL 360
QY 361 STLAADVYPVDKVSVCYFDDGAAMLSESLAETSEFARKWVPCKKYSIEPRABEMVFAAK 420
Db 361 STLAADVYPVDKVSVCYFDDGAAMLSESLAETSEFARKWVPCKKYSIEPRABEMVFAAK 420
QY 421 IDYLKDKVOTSEYKDRRAKREYEERKIRINMLVSKALCPREGWVMODGTPWPGNNTGD 480
Db 421 IDYLKDKVOTSEYKDRRAKREYEERKIRINMLVSKALCPREGWVMODGTPWPGNNTGD 480
QY 481 HPGMIOVFLGONGGDAEGNELPRLVYVSREKRPQOHKKAGAMALRVSAVLNNGPE 540
Db 481 HPGMIOVFLGONGGDAEGNELPRLVYVSREKRPQOHKKAGAMALRVSAVLNNGPE 540
QY 541 IINLDCDHYIINNSKALREAMCFLMDPNLKGVCYVOPORFDGIDKNDRYANRNTVEFDI 600
Db 541 IINLDCDHYIINNSKALREAMCFLMDPNLKGVCYVOPORFDGIDKNDRYANRNTVEFDI 600
QY 601 NLRGIDGIGPYVYGCVFNRTALYGPPIKVKHKKRSLSKLGGSSRKNSKAKKES 660
Db 601 NLRGIDGIGPYVYGCVFNRTALYGPPIKVKHKKRSLSKLGGSSRKNSKAKKES 660
QY 661 DKKSGRHTDSTVPVFNLDIEGEGAGFDEKALMOMSLERFGOSAVVASTIME 720
Db 661 DKKSGRHTDSTVPVFNLDIEGEGAGFDEKALMOMSLERFGOSAVVASTIME 720
QY 721 NGGVPSPATPEMLKEAIVHISGVEDKSDWMEIGWISYTEDILLTFKMHARGWRSI 780
Db 721 NGGVPSPATPEMLKEAIVHISGVEDKSDWMEIGWISYTEDILLTFKMHARGWRSI 780
QY 781 YCMPRLPAKGSAPINLSDRLNOVLKMAISVEILFSRHCPIMYGNGSLKFLERRAYVN 840
Db 781 YCMPRLPAKGSAPINLSDRLNOVLKMAISVEILFSRHCPIMYGNGSLKFLERRAYVN 840
QY 841 TTITPITSIPILMYCTLAVALCTNOFIIPQISINIASIMFLSLFISIFATGILEMWSGV 900
Db 841 TTITPITSIPILMYCTLAVALCTNOFIIPQISINIASIMFLSLFISIFATGILEMWSGV 900
QY 901 GIDEMWRNQQFVYIGVSAHLFAVPOGILKVLADITNTFTVYSKASDEGDEAELEYFKN 960
Db 901 GIDEMWRNQQFVYIGVSAHLFAVPOGILKVLADITNTFTVYSKASDEGDEAELEYFKN 960
QY 961 TLLLPPTLLIVNLGVAVGSAVAINSGYOSMGPLEGKLFAPAFWIVLVLPLKGLMR 1020
Db 961 TLLLPPTLLIVNLGVAVGSAVAINSGYOSMGPLEGKLFAPAFWIVLVLPLKGLMR 1020
QY 1021 QNRFTPIVVMVSVLASIFSLMVRIDPPTSRVTGPDILECGINC 1065
Db 1021 QNRFTPIVVMVSVLASIFSLMVRIDPPTSRVTGPDILECGINC 1065

```

RESULT 2
 Q9FHK6 PRELIMINARY: PRT: 1065 AA.
 ID Q9FHK6
 AC Q9FHK6
 DT 01-MAR-2001 (Tremblrel. 16, Created)

```

DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
DB Cellulose synthase catalytic subunit.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=99397451; PubMed=10470850;
RA Kaneo T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX
RT Sequence features of the regions of 1,011,550 bp covered by seventeen
RT pl and TAC clones."
RL DNA Res. 6:183-195(1999).
DR EMBL: AB018111; BAB09693.1; -;
DR InterPro: IPR005150; Cellulose_synth.
DR InterPro: IPR001841; Znf_ring.
DR Pfam: PF03552; Cellulose_synth; 1.
DR SMART: SM00184; RING; 1
SO SEQUENCE 1065 AA; 119682 MW; 3AA4714CE3C4D581 CRC64;

Query Match 99.6%; Score 5654; DB 10; Length 1065;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1062; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSEGEGTAGKPMKNITVPOTCOICSDNVGKTVDGRVADICSFPCRCREYERDNG 60
Db 1 MSEGEGTAGKPMKNITVPOTCOICSDNVGKTVDGRVADICSFPCRCREYERDNG 60
QY 61 SCPOCKTRKRLKGSALIPGDDKDEDLAEGTVEFVYPOKEKISERMILMHLTRGGEEM 120
Db 61 SCPOCKTRKRLKGSALIPGDDKDEDLAEGTVEFVYPOKEKISERMILMHLTRGGEEM 120
QY 121 GEPQYDKESVHNLPLRLTSRODTSGEFSAASPERLSVSTIAGKRLPYSSDNOGPNRR 180
Db 121 GEPQYDKESVHNLPLRLTSRODTSGEFSAASPERLSVSTIAGKRLPYSSDNOGPNRR 180
QY 181 IYDPVGLGNVAMKERVYDGMKMKOEKNTGVSVOAASERGVDIDASTDILADEALLNDEA 240
Db 181 IYDPVGLGNVAMKERVYDGMKMKOEKNTGVSVOAASERGVDIDASTDILADEALLNDEA 240
QY 241 ROPLSKRVSSIPSSRINPYRMVIMLRVYICLFHYRITNPVPAFALMLVSYCEIEMFAL 300
Db 241 ROPLSKRVSSIPSSRINPYRMVIMLRVYICLFHYRITNPVPAFALMLVSYCEIEMFAL 300
QY 301 SWILDOFPKMFVFNRETYIDRLALRYDRGEGESQOLAVIDIFSTVDPLEKPLVTANTYL 360
Db 301 SWILDOFPKMFVFNRETYIDRLALRYDRGEGESQOLAVIDIFSTVDPLEKPLVTANTYL 360
QY 361 STLAADVYPVDKVSVCYFDDGAAMLSESLAETSEFARKWVPCKKYSIEPRABEMVFAAK 420
Db 361 STLAADVYPVDKVSVCYFDDGAAMLSESLAETSEFARKWVPCKKYSIEPRABEMVFAAK 420
QY 421 IDYLKDKVOTSEYKDRRAKREYEERKIRINMLVSKALCPREGWVMODGTPWPGNNTGD 480
Db 421 IDYLKDKVOTSEYKDRRAKREYEERKIRINMLVSKALCPREGWVMODGTPWPGNNTGD 480
QY 481 HPGMIOVFLGONGGDAEGNELPRLVYVSREKRPQOHKKAGAMALRVSAVLNNGPE 540
Db 481 HPGMIOVFLGONGGDAEGNELPRLVYVSREKRPQOHKKAGAMALRVSAVLNNGPE 540
QY 541 IINLDCDHYIINNSKALREAMCFLMDPNLKGVCYVOPORFDGIDKNDRYANRNTVEFDI 600
Db 541 IINLDCDHYIINNSKALREAMCFLMDPNLKGVCYVOPORFDGIDKNDRYANRNTVEFDI 600
QY 601 NLRGIDGIGPYVYGCVFNRTALYGPPIKVKHKKRSLSKLGGSSRKNSKAKKES 660
Db 601 NLRGIDGIGPYVYGCVFNRTALYGPPIKVKHKKRSLSKLGGSSRKNSKAKKES 660

```


QY 661 DKKSGRHTDSTVPVFNLDIEEGVAGFDEKALMSOMSLERFGQSAVFVASTIME 720
DB 661 DKKSGRHTDSTVPVFNLDIEEGVAGFDEKALMSOMSLERFGQSAVFVASTIME 720
QY 721 NGCVPPSTPNNLKEAIVHISCGYEDKSDWMEIGWYGVTEIDILTGFKHARGMSI 780
DB 721 NGCVPPSTPNNLKEAIVHISCGYEDKSDWMEIGWYGVTEIDILTGFKHARGMSI 780
QY 781 YCMKPLPAFKGSAPINLSDRINOVLRMALGSEILFNRHCPHWGYNRKLFLERFAVN 840
DB 781 YCMKPLPAFKGSAPINLSDRINOVLRMALGSEILFNRHCPHWGYNRKLFLERFAVN 840
QY 841 TTIVPITSIPILMYCTLLAVCLFTNOFTIPIOSINIASIWFLSLFLSIFATGILERMMSGV 900
DB 841 TTIVPITSIPILMYCTLLAVCLFTNOFTIPIOSINIASIWFLSLFLSIFATGILERMMSGV 900
QY 901 GIDEMWNEQFWYIGVSAHLFAVFOGILKVLADITNFTVTSKASDEGDGFAELYLEKM 960
DB 901 GIDEMWNEQFWYIGVSAHLFAVFOGILKVLADITNFTVTSKASDEGDGFAELYLEKM 960
QY 961 TLLIIPPTLLIYNVGVVAGSVAINSGYSGWGLFKLFAFVAVIHLVPELGLMGR 1020
DB 961 TLLIIPPTLLIYNVGVVAGSVAINSGYSGWGLFKLFAFVAVIHLVPELGLMGR 1020
QY 1021 QNRPTIYVWVSVLLASIFSLIMVRIIDPFTSRVTGPDIIEGGINC 1065
DB 1021 QNRPTIYVWVSVLLASIFSLIMVRIIDPFTSRVTGPDIIEGGINC 1065

RESULT 3

QY 09XG6 PRELIMINARY: PRT: 1067 AA.
AC 09XG6:
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Cellulose synthase catalytic subunit.
GN CEL3.
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OX eurosids II; Malvales; Malvaceae; Gossypium.
RN NCB1_TaxID=3635;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=CV. TEXAS MARKER-1; TISSUE=COTTON FIBER;
RX MEDLINE=20098737; PubMed=10631273;
RA laosincha1 W.; Cui X.; Brown R.M. Jr.;
RT "A full length cDNA of cotton cellulose synthase has high homology
with the Arabidopsis RSM1 gene and the cotton Cel1 gene (Accession
No. AF200453) (PCR 00-002).";
RL Plant Physiol. 122:291-291(2000).
RN [2]
RN SEQUENCE FROM N. A.
RC STRAIN=CV. TEXAS MARKER-1; TISSUE=COTTON FIBER;
RA Kimura S.; laosincha1 W.; Itoh T.; Cui X.; Brown R.M. Jr.;
RT "Immunogold labeling of Rosette Terminal Cellulose Synthetizing
Complexes in a Vascular Plant (Vigna angulatis).";
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF150630; AAD39534.2; -
DR InterPro: IPR005150; Cellulose_synth.
DR InterPro: IPR001841; znf_ring.
DR Pfam: PF03552; Cellulose_synth; 1.
DR SMART: SM00184; RING; 1.
SQ SEQUENCE 1067 AA; 119325 MW; 0BBA2ED00590F29C CRC64;

Query Match 87.3%; Score 4954; DB 10; Length 1067;
Best Local Similarity 85.3%; Pred. No. 0;
Matches 919; Conservative 75; Mismatches 60; Indels 24; Gaps 6;

QY 1 MESGEAGAKPMKKNIPVOTQICSDNVGKTVGDGRFVACDICSFFVRCPCYERKRDGNO 60
DB 1 MESEGDIGKPMKNIGGQTCICGDNVGNKTDGDFIACNICAFVRCPCYERKRDGNO 60

QY 61 SCPOCKTRKRLKSGPAPLPGDKEDGLADEGVFE-----NYPOKEKISERMLGHTRGK 116
DB 61 SCPOCKTRKRLKSGPAPLPGDKEDGLADEGVFE-----NYPOKEKISERMLGHTRGK 116
QY 117 GEEMGEPOYDKEVSHNHLPRLSRODTSGEFSAAAPERLSVST-IAGGKRLPYSSDVNO 175
DB 121 GEDVCAPTYDKRISHNHIFPLTSGOEVSGELSAAPERLSMAVPGVAGK----- 170
QY 176 SPNRKIVPVY-----GIGNNAKFRVNGKMKKOKKNGPVST-QAASRGVDIDASTD 228
DB 171 -SSIRVDPVREFGSSGIGNNAKFRVNGKMKKOKKNGPVST-QAASRGVDIDASTD 229
QY 229 ILADBLANDEAROPLSKRSVIPSRRIPYRNVILRIYICLFLPHYRTNVVPAFALM 288
DB 230 VLVDDSQLNDEAROPLSKRSVIPSRRIPYRNVILRIYICLFLPHYRTNVVPAFALM 289
QY 289 LVSVLCEIWFALSWILDQFPKWFVNNRETYLDRLALRYDREGEPSQLAAVDIFVSTVDP 348
DB 290 LISVCEIWFALSWILDQFPKWFVNNRETYLDRLALRYDREGEPSQLAAVDIFVSTVDP 349
QY 349 KEPPLVTANTVLSIANDPVNDKVCYFDDGAMLSFESLAEISEFARKVVPCKKYSI 408
DB 350 KEPPLVTANTVLSIANDPVNDKVCYFDDGAMLSFESLAEISEFARKVVPCKKYSI 409
QY 409 EPRAREWYFAKIDYLDKQVOTSEFKDRAMREVEEKKIRINALVSKALCPREGWMO 468
DB 410 EPRAREWYFAKIDYLDKQVOTSEFKDRAMREVEEKKIRINALVSKALCPREGWMO 469
QY 469 DGTWPGNNTGDHPGMIQVFLQNGCLDAEGNELPRLVYVSEKRRPGFHHKAGAMNL 528
DB 470 DGTWPGNNTGDHPGMIQVFLQNGCLDAEGNELPRLVYVSEKRRPGFHHKAGAMNL 529
QY 529 VRVSAVLNTPETLNLDCDHYINNSKALREACFLMDPNLQKQVCYVQPFQFDDIDRND 588
DB 530 VRVSAVLNTPETLNLDCDHYINNSKALREACFLMDPNLQKQVCYVQPFQFDDIDRND 589
QY 589 RYANRNTVEFDNLNLGLDGIQGPVYVGVGVNRRALGYEPRIVKHKKPSLLSKLGG 648
DB 590 RYANRNTVEFDNLNLGLDGIQGPVYVGVGVNRRALGYEPRIVKHKKPSLLSKLGG 649
QY 649 SRKKNKSK-AKRESDDKSGRHTDSTVPVFNLDIEEGVAGFDEKALMSOMSLERF 707
DB 650 SRKKNKSKSKSGKSGKSHVSTVPVFNLDIEEGVAGFDEKALMSOMSLERF 709
QY 708 GQSAVFVASTIMENGVPPSATPENLLEAIVHISCGYEDKSDWMEIGWYGVTEIDIL 767
DB 710 GQSAVFVASTIMENGVPPSATPENLLEAIVHISCGYEDKSDWMEIGWYGVTEIDIL 769
QY 768 TGFKNHARGMSIYCMKPLPAFKGSAPINLSDRINOVLRMALGSEILFNRHCPHWGYN 827
DB 770 TGFKNHARGMSIYCMKPLPAFKGSAPINLSDRINOVLRMALGSEILFNRHCPHWGYN 829
QY 828 GLKFLERFAVYNTIYVITSIPILMYCTLLAVCLFTNOFTIPIOSINIASIWFLSLFLSI 887
DB 830 GLKFLERFAVYNTIYVITSIPILMYCTLLAVCLFTNOFTIPIOSINIASIWFLSLFLSI 889
QY 888 FATGILEMRMSGVIGIDEMWNEQFWYIGVSAHLFAVFOGILKVLADITNFTVTSKASD 947
DB 890 FATGILEMRMSGVIGIDEMWNEQFWYIGVSAHLFAVFOGILKVLADITNFTVTSKASD 949
QY 948 EDGDFAEIYLFKWTLLIPPTLLIYNVGVVAGSVAINSGYSGWGLFKLFAFVAVI 1007
DB 950 EDGDFAEIYLFKWTLLIPPTLLIYNVGVVAGSVAINSGYSGWGLFKLFAFVAVI 1009
QY 1008 VHLVPELGLMGRQNRPTIYVWVSVLLASIFSLIMVRIIDPFTSRVTGPDIIEGGINC 1065
DB 1010 VHLVPELGLMGRQNRPTIYVWVSVLLASIFSLIMVRIIDPFTSRVTGPDIIEGGINC 1067

RESULT 4
QY 09LLI6 PRELIMINARY: PRT: 1077 AA.
ID 09LLI6

AC Q9L16;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Cellulose synthase-4.
 GN CESA-4.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Zea.
 NCBI_TaxID:4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20398328; PubMed-10938350;
 RA Holland N., Holland D., Helentjaris T., Dhugga K.S.,
 RA Xocostle-Cazares B., Delmer D.P.;
 RT "A comparative analysis of the plant cellulose synthase (Cesa) gene
 family.";
 RL Plant Physiol. 123:1313-1324(2000).
 DR EMBL: AF200528; AAF89964.1;
 DR InterPro: IPR005150; Cellulose_synth.
 DR InterPro: IPR001841; Znf_ring.
 DR Pfam: PF03552; Cellulose_synth. 1.
 DR SMART: SM00184; RING; 1.
 SQ SEQUENCE 1077 AA; 120501 MW; 5F960E4AA73E2D6 CRC64;

Query Match 80.5%; Score 4571; DB 10; Length 1077;
 Best Local Similarity 79.4%; Pred. No. 0;
 Matches 861; Conservative 81; Mismatches 112; Indels 30; Gaps 10;

QY 4 EGCTAG-KPMKNIVPTQCISDNKGTVDGDFVACDICSFVPCRCYERKDNOSC 62
 DB 2 EGDADVKSGRGGGVCICGDDGVTAGDVFAACDVCDFPCRCYERKDNOSC 61

QY 63 PCKTTRYKRLKSPALIPGDKDEGLADEGTEFENY-----PCKEKSERMLGMHLTRCK 116
 DB 62 PCKTTRYKRLKSPALIPGDKDEGLADEGTEFENY-----PCKEKSERMLGMHLTRCK 118

QY 117 GEEMGEPOYDK-----EVSHNHLPLRLTSRODTSGEFAASPERLSVSTIAGK 165
 DB 119 SGVGRPKYDSCSEIGLTKYDSCSEIPRGYIPVTSNSQ-ISEIIGASGDHMMSPPTGNICK 177

QY 166 RLPIYSSDVNOSPRLRYDVGICGVAMKERVGMKKOEKNTGPV---STQASERGV- 221
 DB 178 RAFFPY-VNHSPPSRSEFSGSIGNVAMKERVGMKKOEKNTGPV---STQASERGV- 236

QY 222 DIDASDILADEALNDEAROPLSRKVSPSSRIINPYRMVIMRLVILICFLHYRTINPV 281
 DB 237 DIDASDILADEALNDEAROPLSRKVSPSSRIINPYRMVIMRLVILICFLHYRTINPV 296

QY 282 PNAFALMLVSVICELIFALSWILDOPPKFPPVRETYLDRLALRYDREGESOLAVIDF 341
 DB 297 RNAYPMLLSVICELIFALSWILDOPPKFPPVRETYLDRLALRYDREGESOLAVIDF 356

QY 342 VSVVDLKLPRPLVATNTVSIILAVDVPDKVSCYVDGDAAMLSFSLASTSEPAKWP 401
 DB 357 VSVVDLKLPRPLVATNTVSIILAVDVPDKVSCYVDGDAAMLSFSLASTSEPAKWP 416

QY 402 FCKKYSIEPRAPEMVYAAKIDYLDKRYQTSFVDRAMKREYEEFIRIALVSKLKPC 461
 DB 417 FCKKYSIEPRAPEMVYAAKIDYLDKRYQTSFVDRAMKREYEEFIRIALVSKLKPC 476

QY 462 EBGVWQDSTPWEGNNTGHPGMIQVFLQNGSLADEGMLPRLVYVRSREKRGFQHNK 521
 DB 477 EBGVWQDSTPWEGNNTGHPGMIQVFLQNGSLADEGMLPRLVYVRSREKRGFQHNK 536

QY 522 AGAMNALVRSALVTNGPFLINDCDHYINNSKALBEACFLMDPVLGKOVCYVQRPOR 581
 DB 537 AGAMNALVRSALVTNGPFLINDCDHYINNSKALBEACFLMDPVLGKOVCYVQRPOR 596

QY 582 DGLDKNDRYANRRTVFEDINLRGLDGIQGVYVGTGCVNRTALVYERPIYKHKHPSL 641
 DB 597 DGLDKNDRYANRRTVFEDINLRGLDGIQGVYVGTGCVNRTALVYERPIYKHKHPSL 654

QY 642 LSKLCGSRKRNKAKKESDGRHTDSTVEFVFNLDIEEGVAGFEDERALLMSQ 701
 DB 655 LSSLCGG-RKKAKSKKSGSKKSKQKHVDSVVFNLEDIEEGVAGFEDERALLMSQ 713

QY 702 SLEKRRGQSAVPAVSTLMENGVPSPATPENLKEAIVHISGCEYEDKSDMGMTIGWYGS 761
 DB 714 SLEKRRGQSAVPAVSTLMENGVPSPATPENLKEAIVHISGCEYEDKSDMGMTIGWYGS 773

QY 762 VTEEDILTFEPMHARGMSIYCMKPLAFKASAPINISDRNLQVLRNALGSVEILFSRHC 821
 DB 774 VTEEDILTFEPMHARGMSIYCMKPLAFKASAPINISDRNLQVLRNALGSVEILFSRHC 833

QY 822 IWTYNGRLKLELREFAVNTTIPITISIPILMCTLLAVCLFTNOFTIPOISIASIWL 881
 DB 834 IWTYNGRLKLELREFAVNTTIPITISIPILMCTLLAVCLFTNOFTIPOISIASIWL 893

QY 882 SLFLSIFANGILEMRSGVCIDEMWRNEQFWIGVSAHLFAVQGLKVLADINFTY 941
 DB 894 SLFLSIFANGILEMRSGVCIDEMWRNEQFWIGVSAHLFAVQGLKVLADINFTY 953

QY 942 TBKASDEGDFAEIYFKWTTLLIPPTLLIVNLGVVAGVSAINSQSWEPRLRLE 1001
 DB 954 TBKASDEGDFAEIYFKWTTLLIPPTLLIVNLGVVAGVSAINSQSWEPRLRLE 1013

QY 1002 FAFWVIVHLVPLFKLGMGRONRPTIVVWSVYLASFSLVWRIIDPFSRYVGPDLRC 1061
 DB 1014 FAFWVIVHLVPLFKLGMGRONRPTIVVWSVYLASFSLVWRIIDPFSRYVGPDLRC 1073

QY 1062 GINC 1065.
 DB 1074 GINC 1077

RESULT 5
 ID Q9L16 PRELIMINARY; PRT: 1079 AA.
 AC Q9L16;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Cellulose synthase-9.
 GN CESA-9.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Zea.
 NCBI_TaxID:4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20398328; PubMed-10938350;
 RA Holland N., Holland D., Helentjaris T., Dhugga K.S.,
 RA Xocostle-Cazares B., Delmer D.P.;
 RT "A comparative analysis of the plant cellulose synthase (Cesa) gene
 family.";
 RL Plant Physiol. 123:1313-1324(2000).
 DR EMBL: AF200533; AAF89969.1;
 DR InterPro: IPR005150; Cellulose_synth.
 DR InterPro: IPR001841; Znf_ring.
 DR Pfam: PF03552; Cellulose_synth. 1.
 DR SMART: SM00184; RING; 1.
 SQ SEQUENCE 1079 AA; 120762 MW; EB34272C2ED2D07 CRC64;

Query Match 80.4%; Score 4564; DB 10; Length 1079;
 Best Local Similarity 79.1%; Pred. No. 0;
 Matches 858; Conservative 86; Mismatches 111; Indels 30; Gaps 11;

QY 4 EGCTAG-KPMKNIVPTQCISDNKGTVDGDFVACDICSFVPCRCYERKDNOSC 62
 DB 2 EGDADVKSGRGGGVCICGDDGVTAGDVFAACDVCDFPCRCYERKDNOSC 61

QY 63 PCKTTRYKRLKSPALIPGDKDEGLADEGTEFENY-----PCKEKSERMLGMHLTRCK 116
 DB 62 PCKTTRYKRLKSPALIPGDKDEGLADEGTEFENY-----PCKEKSERMLGMHLTRCK 118

```

Db      62  PCKNNKYYKHKSGPAIRGEEDDTDADAS--DFNYPASGNDQOKIADRMKSWRMNAG 120
Oy      117  GEEMGEPOYDK-----EVSHNHLRLTSPRODTSGEPSAASPERLSVSTIAGK 165
Db      121  SGDVRPKYIDSGEITLTYDGEITPRGITPSYTNQ--ISGELPGASPDHMHMSPGNTGR 179
Oy      166  RLPS--SDVNSPNRRIYDPVGLGNVAMKERYDGMKMKOEKNTGVP---STQASERGV 221
Db      180  RAPEFYMHMSSNPSEFGSV--GNVAMKERDGMKMKODKKTITMTNTSTAPSGRY 237
Oy      222  -DIDASTDLADEALLNDEAROPLSRKVSIPSSRINPYRMVIMRLVILCLFLHYRTNP 280
Db      238  GDIDASTDYNMEDALLNDETROPLSRKVPPLPSSRINPYRMVIVLRLIVLSTFLHYRTNP 297
Oy      281  VPAFALMLVSYICETMFALSWILDOFPKMPVNETYDRLAYDRGEPSOLAADI 340
Db      298  VRNAPPLMLVSYICETMFALSWILDOFPKMPVNETYDRLALRDREGEPSOLAADI 357
Oy      341  FVSTVDPLKEPPLVANTVLSILAVDYPVKVSCYVDDGAMLSPESLAETSEFARKV 400
Db      358  FVSTVDPMKEPPLVANTVLSILAVDYPVKVSCYVDDGAMLTFDALAETSEFARKV 417
Oy      401  PECKKSYIEPRAPERYFAAKIDYLDKQVQTSVVDKRRAMKREYEEFKIRINALVSKALKC 460
Db      418  PEVKRYNIEPRAPERYFSQIDYLDKDKVHPSEVDKRRAMKREYEEFKIRYNGLVAKQKY 477
Oy      461  PEEGVNMDGTPWPNNTGDHPGMTIOVFLGONGDAEBNELPRLVYVSREKRPQOHK 520
Db      478  PEEGVNMDGTPWPNNTGDHPGMTIOVFLGSHSGIDTEBNEPLRLVYVSREKRPQOHK 537
Oy      531  KAGANNALVRSVAVLTNGPFILNDCDHYINNSKALREAMCFLMPNLKQVCYQVOPOR 580
Db      538  KAGANNALVRSVAVLTNGQYMLNDCDHYINNSKALREAMCFLMPNLGRSCYQVOPOR 597
Oy      561  FDGIDKNDRIANRNTVFEDINLRGLDIOGPVYVGTGCVFNRTALYGPPIKVKHKPS 640
Db      598  FDGIDRNDRIANRNTVFEDINLRGLDIOGPVYVGTGCVFNRTALYGPPIKVKHKPS 655
Oy      641  LLSKLCGSRKKNSKAKKESDCKKGRHNDSTVPPNLDIDIEGVEGAGFDEKSLMSQ 700
Db      656  FLSSLCGG-RKKGSKSKGSDCKKQKQKHYDSSVPVFNEDIEGVEGAGFDEKSLMSQ 714
Oy      701  MSLERFGQSAFVASTLMENGVPPSATPENILKEAIIHVISGCEYEDKSDMCMETIGW 760
Db      715  MSLERFGQSAFVASTLMETGVQSAATPESILKEAIIHVISGCEYEDKTEMETIGW 774
Oy      761  SVTEDILITGFKMHAGMSIYCMKPLPAFKGSAPINLSDRLNOVLKMGALGVEILLFSRH 820
Db      775  SVTEDILITGFKMHAGMSIYCMKPRPAFKGSAPINLSDRLNOVLKMGALGVEILLFSRH 834
Oy      821  PIWGYNGRLKLEFPAYVNTTIPYTSIPILMCGTILAVCLFTNOFTIROSINASTMF 880
Db      835  PIWGYNGRLKLEFPAYVNTTIPYTSIPILMCGTILAVCLFTNOFTIROSINASTMF 894
Oy      881  LSLFLSIFATGILEKRWGSGIDEMWRNEQFVIGVSAHLEFAVEGILKVLVAGIDTNT 940
Db      895  LSLFLSIFATGILEKRWGSGIDEMWRNEQFVIGVSAHLEFAVEGILKVLVAGIDTNT 954
Oy      941  VTSKASDEBDGFAELYKWTLLLPPTLLLVNLGVVAVGVSYAINSGYSGMPLFGKL 1000
Db      955  VTSKASDEBDGFAELYKWTLLLPPTLLLVNLGVVAVGVSYAINSGYSGMPLFGKL 1014
Oy      1001  FFAFVVIHLVPELGLMGORRPTIYVVMVSVLLASTFSLMVRIDPSTSVTGDILE 1060
Db      1015  FFAFVVIHLVPELGLMGORRPTIYVVMVSVLLASTFSLMVRIDPSTSVTGDILE 1074
Oy      1061  CGINC 1065
Db      1075  CGINC 1079

```

```

ID      09LIL5  PRELIMINARY;  PRT;  1076 AA.
AC      09LIL5;
DT      01-OCT-2000 (Tremblrel. 15, Created)
DT      01-OCT-2000 (Tremblrel. 15, last sequence update)
DE      01-JUN-2002 (Tremblrel. 21, last annotation update)
DE      Cellulose synthase-5.
GN      CESA-5.
OS      Zea mays (maize).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC      Panicoideae; Andropogoneae; Zea.
OX      NCBI_TaxID=4577;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20398328; PubMed=10938350;
RA      Holland N., Holland D., Helentjaris T., Dhugga K.S.,
RA      Xocoostle-Cazares B., Delmer D.P.;
RT      "A comparative analysis of the plant cellulose synthase (Cesa) gene
RT      family";
RL      Plant Physiol. 123:1313-1324(2000).
DR      EMBL: AF200529; AAF89965.1;
DR      InterPro: IPR005150; Cellulose_synth.
DR      InterPro: IPR001841; Znf_ring.
DR      Pfam: PF03552; Cellulose_synth; 1.
DR      SMART: SM00184; RING; 1.
SQ      SEQUENCE 1076 AA; 120724 MW; 41E7C20EDA54F27 CRC64;

Query Match      79.8%; Score 4529.5; DB 10; Length 1076;
Best Local Similarity 78.1%; Pred. No. 0;
Matches 847; Conservative 88; Mismatches 118; Indels 31; Gaps 10;

Oy      4  EEEETAGKPKKNIVPOTCOICSDNVKTVGDRFVACDISFPYCRPCYERKDNQOSP 63
Db      2  DGGDATNSCKHYAGVQCICGCGVGTADGDLFTACDVGCFYCRPCYERKDKGQACP 61
Oy      64  OCTKRYRLKGSPIAPGDEGLADEGVEFNY-----POKEKISEBMLNHLTRGK 117
Db      62  OCTKRYKHKSGSPVHGEENEVDVADD--VSDNYQASGNODQOKIAEMLTMR--TNSG 119
Oy      118  EEMGEPOYDK-----EVSHNHLRLTSPRODTSGEPSAASPERL--SVSSTIAG 164
Db      120  SDIGLAKYDSEIGHGKYDSCGELPRNGYISLTHSQ--ISGELPGASPDHMHMSPGNTGR 178
Oy      165  KRLPSYSDVNSPNRRIYDPVGLGNVAMKERYDGMKMKOGAIIPMTNGSTIAPSGRY 221
Db      179  HQPPT--VNHSNPSRERFSGSLGNVAMKERYDGMKMKOGAIIPMTNGSTIAPSGRY 235
Oy      222  DIDASTDLADEALLNDEAROPLSRKVSIPSSRINPYRMVIMRLVILCLFLHYRTNP 281
Db      236  DIDASTDYNMEDALLNDETROPLSRKVPPLPSSRINPYRMVIVLRLIVLSTFLHYRTNP 295
Oy      282  PNAFALMLVSYICETMFALSWILDOFPKMPVNETYDRLALRYDRGEPSOLAADI 341
Db      296  PNAFALMLVSYICETMFALSWILDOFPKMPVNETYDRLALRYDRGEPSOLAADI 355
Oy      342  VSTVDPLKEPPLVANTVLSILAVDYPVKVSCYVDDGAMLSPESLAETSEFARKV 401
Db      356  VSTVDPMKEPPLVANTVLSILAVDYPVKVSCYVDDGAMLTFDALAETSEFARKV 415
Oy      402  PECKKSYIEPRAPERYFAAKIDYLDKQVQTSVVDKRRAMKREYEEFKIRINALVSKALKC 461
Db      416  PECKKSYIEPRAPERYFAAKIDYLDKQVQTSVVDKRRAMKREYEEFKIRINALVSKALKC 475
Oy      462  EEEGVNMDGTPWPNNTGDHPGMTIOVFLGONGDAEBNELPRLVYVSREKRPQOHK 521
Db      476  EEEGVNMDGTPWPNNTGDHPGMTIOVFLGSHSGIDTEBNEPLRLVYVSREKRPQOHK 535
Oy      522  KAGANNALVRSVAVLTNGPFILNDCDHYINNSKALREAMCFLMPNLKQVCYQVOPOR 581
Db      536  KAGANNALVRSVAVLTNGQYMLNDCDHYINNSKALREAMCFLMPNLGRSCYQVOPOR 595
Oy      582  FDGIDKNDRIANRNTVFEDINLRGLDIOGPVYVGTGCVFNRTALYGPPIKVKHKPS 641

```

Db 596 DGIIDRRDRIANRNTVFEDINLRLGLDGIQGFVYVTCGVENRTALYGEPP--VKKKKPGF 653
QY 642 LSKLCGSGRRKNSKAKKESDCKSGRHTDSTVPFNLDIEEGVEGAGFDEKALLMSQM 701
Db 654 FSSLCGG-RKTKSKSKSKSKSHRHAOSVFPNLEDIEEGISQGFDEKSLMSQM 712
QY 702 SLEKRRGQSAVFPASTLMEKGVPPSPATPENLKEAIVHISCGYEDKSDMGMEIGWYGS 761
Db 713 SLEKRGQSSVFPASTLMEKGVPPSPATPENLKEAIVHISCGYEDKSDMGMEIGWYGS 772
QY 762 VTEIDITGFPMHARGRSIYCMPLAFAGSAPINISDRINQVLRNALGSVELLESHCP 821
Db 773 VTEIDITGFPMHARGRSIYCMPLAFAGSAPINISDRINQVLRNALGSVELLESHCP 832
QY 822 IMVNGRLKFLERFAYVNTIYPTISIPLLMYCTLLAVCLFTNOFIPOISMIASIMFL 881
Db 833 IMVNGRLKFLERFAYVNTIYPTISIPLLMYCTLLAVCLFTNOFIPOISMIASIMFL 892
QY 882 SLEFLSPATGILEMRSGVCIDEMWRNEQFVIGVSAHLFAVQGLKVLACIDITNFTV 941
Db 893 SLEFLSPATGILEMRSGVCIDEMWRNEQFVIGVSAHLFAVQGLKVLACIDITNFTV 952
QY 942 TSASQEDDGEAEFLYEFKMTTLLIPPTLLIIVNLGVVAGVSAINSQVSGWPLFGKLF 1001
Db 953 TSASQEDDGEAEFLYEFKMTTLLIPPTLLIIVNLGVVAGVSAINSQVSGWPLFGKLF 1012
QY 1002 FAFVAVLHLPFLKMGKRONRPTIYVWWSVLLASIFSLMVRIDPETSRTVGPDLLEC 1061
Db 1013 FAFVAVLHLPFLKMGKRONRPTIYVWWSVLLASIFSLMVRIDPETSRTVGPDLLEC 1072
QY 1062 GINC 1065
Db 1073 GINC 1076

RESULT 7
09LL19 PRELIMINARY; PRT; 1075 AA.
ID 09LL19; PRT; 1075 AA.
AC 09LL19;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Cellulose synthase-1.
GN CESA-1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_taxid=4577;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20398328; PubMed-10938350;
RA Holland N., Holland D., Helentjaris T., Dhugra K.S.,
RA Xocostole-Cazares B., Delmer D.P.,
RT A comparative analysis of the plant cellulose synthase (Cesa) gene
family.
RL Plant Physiol. 123:1313-1324(2000).
DR EMBL: AF200525; AAF89961.1; .
DR InterPro: IPR005150; Cellulose_synth.
DR InterPro: IPR001841; Znf_ring.
DR Pfam: PF03552; Cellulose_synth. 1.
DR SMART: SM00184; RING. 1.
SQ SEQUENCE 1075 AA; 121181 MW; 67A4FBD97A811F33 CRC64;

Query Match 71.0%; Score 4028.5; DB 10; Length 1075;
Best Local Similarity 69.6%; Pred. No. 4e-304;
Matches 752; Conservative 130; Mismatches 153; Indels 45; Gaps 12;

QY 4 ESETAG--KPMNIVPOTQICSDNVGKTVGDRFVACICSPVPCPYEYRKDGNOS 61
Db 23 DDDVPSAKPTKSANGVQICGDSVGSATGTVFACMCACAPVPCPYEYRKDGNOC 82
QY 62 CPCKTRRYRLKAGSPALPDCKDEGLADGVEFENYPOKESIRMLGMHLTRGKEEMG 121

Db 83 CPCKTRRYRKOKSPRVHGEDEEDV--DDLDNFEFNKOGS-----GKGPENW 128
QY 122 EPQYDEVS-----HNHLPRLTRODTSGEESAPESLVSSTIAGKRIPYSVDN 174
Db 129 LOGDDADLSSAAHEPHRIRLRLTSGQOISGELPDSAPDRHSI-----RPTSSYND 180
QY 175 QS--PNRRIVDP-----VGLGNVANKERYDGMKMKOEKNTGPTOASERGVNDAS 226
Db 181 PSYVPRVRIYDPSKDLNYSYGLNSVDKEREVSRRVQKDKMMQVTKNYPARGG--DME-G 238
QY 227 TDLADEALLNDAROPLSKVSIPSSRIIPYVIMLRLVITCLELHRITIPYRNFA 286
Db 239 TGSNGEMQAVDDARLPLSRIVPISNQNLVYVVIILRLILICFFQYRVSHPVDAYG 298
QY 287 LMLVSYCEIWMFLSMTLDQFPKMPVNRETYLDRLALRDREGEESQLAVIDPSTVD 346
Db 299 LMLVSYCEIWMFLSMTLDQFPKMPVNRETYLDRLALRDREGEESQLAVIDPSTVD 358
QY 347 PLKEPPLVANTVLAIVADYVDYKVCYVFDGGAAMLSESLAETSEPAKRVPPCKRY 406
Db 359 PLKEPPLVANTVLAIVADYVDYKVCYVFDGGAAMLSESLAETSEPAKRVPPCKRY 418
QY 407 STEPRAPWYFAAKIDYLDKQVQTSYKDRRAKREYEEKIRINLVSKALKCPDEGW 466
Db 419 NIEPRAPWYFAAKIDYLDKQVQTSYKDRRAKREYEEKIRINLVSKALKCPDEGW 478
QY 467 MDSCTPMPGNNPDHGMIOVFLGONGGDAENELPRLYVSRERKPGOHKKKAGAM 526
Db 479 MADGTAMPGNNPDHGMIOVFLGONGGDAENELPRLYVSRERKPGOHKKKAGAM 538
QY 527 ALTVASAVLNGPFIILNCDHYINNSKALREAMECELMNPNGKQVYVQFPRFGIDK 586
Db 539 ALTVASAVLNGPFIILNCDHYINNSKALREAMECELMNPNGKQVYVQFPRFGIDK 598
QY 587 NDRYANRNTVFEDINLRLGLDGIQGFVYVTCGVENRTALYGEPP--VKKKKPGF 653
Db 599 NDRYANRNTVFEDINLRLGLDGIQGFVYVTCGVENRTALYGEPP--VKKKKPGF 666
QY 647 CGSRKKNSKAKKESDCKSGRHTDSTVPFNLDIEEGVEGAGFDEKALLMSQM 701
Db 659 CGSRKKNSKAKKESDCKSGRHTDSTVPFNLDIEEGVEGAGFDEKALLMSQM 712
QY 707 FGQSAVFVASTLMEKGVPPSPATPENLKEAIVHISCGYEDKSDMGMEIGWYGS 766
Db 716 FGQSAVFVASTLMEKGVPPSPATPENLKEAIVHISCGYEDKSDMGMEIGWYGS 772
QY 767 LTGFKMHARGMSIYCMPLAFAGSAPINISDRINQVLRNALGSVELLESHCP 821
Db 776 LTGFKMHARGMSIYCMPLAFAGSAPINISDRINQVLRNALGSVELLESHCP 835
QY 827 NGRLKFLERFAYVNTIYPTISIPLLMYCTLLAVCLFTNOFIPOISMIASIMFL 881
Db 836 NGRLKFLERFAYVNTIYPTISIPLLMYCTLLAVCLFTNOFIPOISMIASIMFL 892
QY 887 IFATGILEMRSGVCIDEMWRNEQFVIGVSAHLFAVQGLKVLACIDITNFTV 941
Db 896 IFATGILEMRSGVCIDEMWRNEQFVIGVSAHLFAVQGLKVLACIDITNFTV 952
QY 947 DEDGDFAEFLYEFKMTTLLIPPTLLIIVNLGVVAGVSAINSQVSGWPLFGKLF 1001
Db 956 DEDGDFAEFLYEFKMTTLLIPPTLLIIVNLGVVAGVSAINSQVSGWPLFGKLF 1012
QY 1007 IYHLYPFLKMGKRONRPTIYVWWSVLLASIFSLMVRIDPETSRTVGPDLLEC 1061
Db 1016 IYHLYPFLKMGKRONRPTIYVWWSVLLASIFSLMVRIDPETSRTVGPDLLEC 1072

RESULT 8
09LL18 PRELIMINARY; PRT; 1074 AA.
ID 09LL18;
AC 09LL18;
DT 01-OCT-2000 (Tremblrel. 15, Created)

DT	01-OCT-2000 (TREMBlrel. 15, last sequence update)
DT	01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE	Cellulose synthase-2.
GN	CESA-2.
OS	Zea mays (Maize).
CC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
CC	Panicoidae; Andropogoneae; Zea.
OX	NCBI_TaxID:4577;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE:20398328; PubMed:10938350;
RA	Holland N., Holland D., Helentjaris T., Dhingra K.S.,
RA	Xocoonstle-Cazares B., Delmer D.P.;
RT	"A comparative analysis of the plant cellulose synthase (Cesa) gene
RT	family.";
RL	Plant Physiol. 123:1313-1324(2000).
DR	EMBL: AF200526; AAF69962.1; -
DR	InterPro: IPR005150; Cellulose_synth.
DR	InterPro: IPR001841; Znt_ring.
DR	Pfam: PF03552; Cellulose_synth; 1.
DR	SMART: SM00184; RING; 1.
SO	SEQUENCE 1074 AA; 121187 MW; A11666F0564E210 CRC64;
Query Match	70.9%; Score 4024; DB 10; Length 1074;
Best Local Similarity	69.4%; Pred. No. 9e-304;
Matches 749; Conservative 128; Mismatches 159; Indels 44; Gaps 10;	
OY	2 ESEGETAGKPKNKINIVPOTCOICSDNVGKTVGDRPVACDICSFPCRCPCYEERKDGNS 61
DB	23 DGDAPVPAPKPKTSANGVCQICGDTVGATGDVFAACNECAFVPCRCPCYEERKGNOC 82
OY	62 CPCKTRKRYRLKSGSPALFPGDDEGLADEGVFEFNPQEKISEBMLGMHLTRGGEEMG 121
DB	83 CPCKTRKRYRKSGSPRVHGDDDEEDV-DLDLNEFNYKQ-----GNGKGPMEQ 128
OY	122 EPQYDKES-----HNHLRLTISRQDTSGEFSASPERLSVSTIAGKRRLPYSDVN 174
DB	129 LQGDADADLSSASARHDPHRIPLRTLSGQDISGEIPLASPDSRHSI-----RSPSSSYVD 180
OY	175 QS--PNRRIVP-----VGLGNVAMKRRVGVGMKKOKKNGPVSQTQASERGVSDIDAS 226
DB	181 PSVPVPARIVPSSDLNSTYGLNSVDMKRRVESRWVKQKQNMLOVYNNKPEARGDWE---G 237
OY	227 TDILADEALLDEARQPLSRKVSIPSSRINPVMYIMDLRLVILFLHYRTINPVPNFA 286
DB	238 TGSNGEDQMWDARLPISRIVISSNOLNLXRIYIILRLIICFFPQYRISHPRNAYG 297
OY	287 LMLVSVICELFALSWILDOFPKMEPVNRETYLDRLALRYDREGSPQLAAVDIFVSTD 346
DB	298 LMLVSVICELFALSWILDOFPKMPYINRETYLDRLALRYDREGSPQLAPDIVEVSTD 357
OY	347 PLKRPPLVTAMTVLSILAVDYVPVKVSGCYVDDGAMLSFESLATSSEFARKWPFCKKY 406
DB	358 PLKRPPLVTAMTVLSILAVDYVPDVSVSCYSDDGAMLTFSLSSTAEFAKRWPFCKKH 417
OY	407 SIEPRAPEMYAAKIDLYLKDQVQTSFVDRKAMKREYEEFKRIALVSKALCSKEEGV 466
DB	418 NIEPRAPERYAQKIDLYLKDQIOPSFAERARMKREYEEFKRIALVLAQAKQVYEEGVT 477
OY	467 MODGTPMEGNNTGDHGMIVQVFLGONGSLADEGNELPRLVUYVSREKRPFGOHKKAGAMN 526
DB	478 MADGTAMPGMNDRPHPGMIVQVFLGSGSLDTPDGNELPRLVUYVSREKRPFGOHKKAGAMN 537
OY	527 ALVAVSANVLNTPFLNDCOHYINNSALALEAKMGLMDPNLKGQVSCYVQFRRDGDIDK 586
DB	538 ALIRASAVLTGAVALLNVDCOHYFNSSKALREAKMFMDPALGRKTCYVQFRRDGDIDL 597
OY	587 NDRANRNTVFEEIDLRLDGIQGVYVYGTGCVFRTALVLYGERPIKVKHKKPSLISLTK 646
DB	598 HDRANRNTVFEEIDLRLDGIQGVYVYGTGCVCFKRALVLYGERPIKVKHKKPSLISLTK 657
OY	647 GGSRRKNSKAKESDRKKSGRRHTSDTVFNLMDIEEGVEGAGFDEKALLMSQMSLEKR 706

Db 658 CGRRKKRKNKSMDSQSIIMKR--TESSAPIFMEDIIEGIC--GIEDERSVULMSORAKLEKR 714

Qy 707 FGGSAVEAVASTLMENGCVPPSAPPENILKEAIIHISCGYEDKSDMGMEIGNIGSVTEDI 766
715 FGOSPIFIASPFMIOGGIIPSTNPASLLKEAIHVISCGYEDKTEMKGEIGNIGSVTEDI 774

Qy 767 LTGFKMHARGMRSTICYMPKLPAFGSNPINSRLNOYLNRALGASVELFSRRCPIYGY 826
775 LTGFKMARGOSIYCMPPRCEFGSAPINLSDLNOVLNRALGASVELFSRHCPIMWGY 834

Qy 827 NGRKLFLERFAYVTITPTITSPLMWCTLANCETPNOFIPOINIASIMEFLS 886
835 NGRKLLERLAYINTVIYPTSVPILACVALCLLTNFELPEISNYGMFTILLPAS 894

Qy 887 IFATGILEMRNSGVGIDEMWRNEQFWNIGCVSAHLFAVOGILKVLAGIDTNFTVTSKAS 946
895 IFAGGIELERNSGVGIDEMWRNEQFWNIGCVSAHLFAVOGLKVLVAGIDTNFTVTSKAS 954

Qy 947 DEDGDFAELYLFKWTTLIDPTTLLIVNLGVVAGSYAINSGYSQSGPLFGKLPFAFW 1006
955 DEDGDFAELYLFKWTSLIPTTVLVNLGVVAGSYAINSGYSQSGPLFGKLFESIW 1014

Qy 1007 IVHIYPFLKGLMGQNTPRTIVVWSVLASFESLWVRIDPFPSRYTGPIIL-EGCINC 1055
1015 ILHLYPFLKGLMGQNTPRTIVVWSVLASFESLWVKIDPFISPTOKAAALQGCSVNC 1074

Db Db

RESULT 9

O48946 PRELIMINARY; PRT; 1081 AA.

AC O48946:
DT 01-JUN-1998 (TREMBLrel_06, created)
DT 01-JUN-1998 (TREMBLrel_06, last sequence update)
DT 01-JUN-2002 (TREMBLrel_21, last annotation update)
DE Cellulose synthase catalytic subunit.
GN RSW1 OR FB84.110 OR AT4G32410.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosoids II; Brassicales; Brassicaceae; Arabidopsids.
CX NCBI_TaxId=3702;

RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV, COLUMBIA;
RX MEDLINE=98111412; PubMed=9445479;
RA Ariotti T., Peng L., Betzner A.S., Burn J., Wittke W., Herth W.,
RA Camilleri C., Horte H., Plazinski J., Birch R., Cork A., Glover J.,
RA Redmond J., Williamson R.E.;
RT "Molecular analysis of cellulose biosynthesis in Arabidopsids.";
RL Science 279:717-720(1998).

RN [2]
RP SEQUENCE FROM N.A.
RA Bevan M., Terryn N., Ardiles W., Buysshaert C., Dassevillle R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA Villarroel R., Gelen J., Van Montagu M., Hohelsel J., Mewes H.W.,
RA Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.

RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.

RN [4]
RP SEQUENCE FROM N.A.
RA Terryn N., Ardiles W., Buysshaert C., Dassevillle R., De Clerck R.,
RA De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villaroel R.,
RA Gelen J., Van Montagu M., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.

RN [5]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.

RR EMBL: AF027172; AAC39334.1; -

	Query Match	70.2%	Score 3984;	DB 10;	Length 1081;
	Best Local Similarity	69.7%	Pred. No. 1.2e-300;		
	Matches 741;	Conservative 128;	Mismatches 158;	Indels 36;	Gaps 13;
Dr	EMBL: AL034567: CA422568.1: -				
Dr	EMBL: AL161581: CAB79958.1: -				
Dr	InterPro: IPR005150: Cellulose_synth.				
Dr	InterPro: IPR001841: ZnF_ring.				
Dr	Pfam: PF03552: Cellulose_synth. 1.				
Dr	SMART: SM00184: RING. 1.				
Dr	SEQUENCE 1081 AA; 122236 MW; BDB5D9DE334D59 CRC64;				
Qy	2 ESEGTACKPKMKNVLPQTCOICSDNVGTYVDGDFRVADDCISFVPCRCPCYEERKDDGNS				61
Dd	23 ESDGCT-KPLNNMGQICQICGDDVGAETGDFVAVNCECAFPVCPRCPCYEERKDDGTC				80
Qy	62 CPCKCTRYKRLKGSFAIRPDDEGDEGADGTEGVFNYPCKEKISEFMYLHMLHTRKGGEWG				121
Dd	81 CPCKCTRRRRRRGSRVSGDEDEDV-DDIEENEFYQA-----GANKAHQHRHG				128
Qy	122 EPQYDKEYSHNH--LPRLTSRQDTSGEFSAASPERLSVST--TAGSKRLPYSS--DV				173
Dd	129 E-ESSSSSRHSSQPIPLTLHGHVTSGET--RTPDQTSVRTSGPGLGPDNRMAISSPYID				185
Qy	174 NOSNRRIVDP-----VGLGNVAMKEEVDDGMKMGQENKGTGVSVOAASERGVDDIDAST				227
Dd	186 RQPVPRVRLVSDSKDINSYGLGNVDVKKEREVEGKTLQERNNMLQMTKYEHGKG-EIE-GT				243
Qy	228 DILDEALLNDEARQPLSRKYSISSRIINPYRMVIMLRYILCLFLHRTIPVYNAPAL				287
Dd	244 GSNGBEQLQADMDTRLPMKRVVPISSRLPLRYVILLLILILCFLQYRTHHPVKNAPL				303
Qy	288 WLVSVCIEIWFALSHLDQPFKWPVNRFTYLDRLALRYDREGEPSQLAADVIVSYDP				347
Dd	304 WLTSVCEIWFASFSLDLQFPKWPVIRNETYLDRLAIYDRDGEPSQLVADVFYSTVP				363
Qy	348 LKEPPLVANTVLSLADVVDVSKVCPVDPDGAAMLSFEESIAETSEFARKWVPCCKKYS				407
Dd	364 LKEPPLVANTVLSLSDYVDVADKACVSDGSSMLTFSEISTEFAEPAKKWVPCCKKEN				423
Qy	408 IEPRAPENYFFAKIDYLDKQVQTSFVKKRRAMKREYEEFKIRINLVSKALCKPREGVWM				467
Dd	424 IEPRAPENYFFAQKIDYLDKQIQSPVKKRRAMKREYEEFKYRINALVAKQKIPREGWTM				483
Qy	468 ODGTPWPGNNTGDHPGMQVFLGQNGGLDAEGNELPRLVYVSRERKPGFOHHKKAGAMA				527
Dd	484 ODGTPWPGNNTRDHPGMQVFLGSHGGLDTPGNEPLRLIYVSRERKPGFOHHKKAGAMA				543
Qy	528 LVRSAYVLTNPFLLNDDCHYINNSKALREAMCTLMPNGLQKQVCYQVQRFQRLGIDKN				587
Dd	544 LVRSAYVLTNAYVLTNDDCHYFNNSKAIKRAMCMMDPAIGKCCYQVQRFQRFEGIDLH				603
Qy	588 DRYANRNNTVFEDIMLRGLDGIQGPVYVGTGCVFNLTALYGEPPYKTVKHKRPSLSKLCG				647
Dd	604 DRYANRNNTVFEDIMMKGLDGIQGPVYVGTGCVFNQALYGYDPTVLTEDLEPNITVSKCC				663
Qy	648 GSRKNSKAKKESDKKSG-RHTDSTVPEVNLDDIEEGVEGAGFDEKALLMSQMSLEKR				706
Dd	664 GSRKKSCKSKKYNEKRGIRNSDSNAPLFNMEDIDEQFE--GYDDERSILMSQSVSEKR				721
Qy	707 FGQSAVPAVSTLMEGCVSPATSEBNLKEAIVHYSQGYEDKSDMGMEIGMYGVTEDI				766
Dd	722 FGQSPVFVLAATFMEQGGIPPTNPAATLLKEAIVHYSQGYEDKTEWKGKIGMYGVTEDI				781
Qy	767 LTGFKMAHARGRSIYCMPLPAFGSAPAINISDRNLQVLRNALSVELTFSHHCITWGY				826
Dd	782 LTGFKMAHARGIISTYCNPPRPAFGSAPAINISDRNLQVLRNALSVELTILSHHCITWGY				841
Qy	827 NGRLLTEREYAVNTTITPITISPLMCTLLAACLTFNOFTIIPDISINIASIETSLFS				886
Dd	842 HGRLLTERIAIVNTIYVPIISPLAACILPAFCILIDRPIIPDISINYSAGIETSLFIS				901
Qy	887 IFATGILBMRWSGVGIDEMWRNEQFWIGVSAHLFAVQGLKVLGADITNFTVTSKAS				946

[illegible]

OY	422	YLMDKQVTSYKQDRAKREVEEKIRIINALVSAKLCRESGVMWQOSTPMRGNNTGCH	481
Db	407	YLMDKQAPNVEPRRAKREVEEKVRNALVAQAQKRESGWMTQDSTPMRGNNVHR	466
OY	482	PGMIQVFLGQNGDLAEAGNELPRILVYVRSERKPRGFOHNHKAAMNALVSVAVLTGSPET	541
Db	467	PGMIQVFLGQSGGHNVEGNELPRILVYVSEKPRGTNHNHKAAMNALVSVAVLTNAPRL	526
OY	542	LNLDCDHYINNSKALREACFLMDPNLGKQCVYQFORFDGIDKNDRYANRNVFEFDIN	601
Db	527	LNLDCDHYINNSKALREACFLMDPNLGKQCVYQFORFDGIDRHDYANRNVFEFDIN	586
OY	602	LRGIDIGQPVYVGVGCVENRATLKGVEPRIVAKRRKS-----LSKLGGSKRK	651
Db	587	MKGIDIGQPVYVGVGCVENRATLKGVEPRIVAKRRKS-----LSKLGGSKRK	643
OY	652	KNSKAKKESDK----KKSGRHTDSYVPVFNLDIDIEGVAGAGFDDKALMSQMSLEKRF	707
Db	644	TKKTKTKTSKPRFEKTKKLEKKENAPVALGEIDEAARGA--ENKASIVMDQKLEKRF	701
OY	708	GQSAAVFASITLMENGVPVPATPENLLKALHIVHISCGYEDKSDMGMEIGMIVGSYTEDIL	767
Db	702	GQSAAVFASITLMENGTLKASAPASLLKEAIVHISCGYEDKTLGWMGKIDIGMIVGSYTEDIL	761
OY	768	TGFKNHAGMSIYMPKLPAPKSGAPINLSRLQVNLWALGSVITLESRCPIYAGIN	827
Db	762	TGFKNHGMSIYCIPIKRAAKGSGAPNLSDFQVNLWALGSIITLESNRCPIYAGY	821
OY	828	GLKLFERFAVNTIYPTISPIPLMYCTGLVLCFTNQFIPTISINIASIMPLSFLSI	887
Db	822	GLKLFERFSTIYPTISPIPLMYCTGLVLCFTNQFIPTISINIASIMPLSFLSIT	881
OY	888	FATGILLENRWSGVGIDEMWRNBOFWVIGGVSAAHFAVFOGILKVLGADITNFTVYSKASD	947
Db	882	FATSILENRWSGVGIDEMWRNBOFWVIGGVSAAHFAVFOGILKVLGADITNFTVYSKASD	941
OY	948	EDGDAEALYTKRWTLILPPTLLLVNLGVYAGVSYAINSQSGKPLFGKLFAPFWYI	1000
Db	942	DEEPEALYTKRWTLILPPTLLLVNLGVYAGVSYAINSQSGKPLFGKLFAPFWYI	1000
OY	1008	VHLVPEFLGLMGORRPTIYVWMSVLASISLWVRIDPFSRYTGPDILIECGINC	1065
Db	1001	VHLVPEFLGLVGRONRPTIYVWMSVLASISLWVRIDPFLAKDGPDLIECGIDC	1058
RESULT 11			
O9LLI3			
ID	O9LLI3	PRELIMINARY:	PRT: 1086 AA.
AC	O9LLI3:		
DT	01-OCT-2000	(TREMBLrel. 15, Created)	
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)	
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)	
DE	Cellulose synthase-7.		
CN	CESA-7.		
OS	Zea mays (Maize).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;		
OC	Panicoideae; Andropogonaceae; Zea.		
OX	NCBI_TaxID=4577;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20398328; PubMed=10938350;		
RA	Holland N., Holland D., Helenjars T., Dhugra K.S.,		
RA	Xoconostle-Cazares B., Delmer D.P.;		
RT	"A comparative analysis of the plant cellulose synthase (cesa) gene		
RT	family.";		
RL	Plant Physiol. 123:1313-1324 (2000).		
DR	EMBL; AF200531; AAF89967.1; -		
DR	InterPro: IPR005150; Cellulose_synth.		
DR	InterPro: IPR001841; znf-ring		
DR	Pfam: PF03552; Cellulose_synth; 1.		
DR	SMART: SM00184; KINg; 1.		
SO	SEQUENCE 1086 AA; 122608 MW; DD03C73ABD13E2EB CRC64;		

Query	Match	Similarity	66.7%	Pred. No.	3.2e-28	DB	Length	1086:
Best	Local	Similarity	66.7%	Pred. No.	3.2e-28	DB	Length	1086:
Matches	728:	Conservative	145:	Mismatches	161:	Indels	57:	Gaps
4	EGTAGPMPKNIYVQTQICSDNVGKVVDDREAVACDICSFPVCRPCYEERKDNQSCP	63						
23	DGDGPKRPBRQNGQVQICGDDVGLAPGDDPVACNECAFPCYCRDCEYERREGTQNC	82						
64	QCKTRRYKRLGSPALIPDKDDEGLADGTYEYVY--QNEKISEBRLGSHHLTRGK-GEEM	120						
83	QCKTRRYKRLKCCQVNTDDEEDGV-DLDLDEFMWDHSDQSAESMLYHMSYGRGDGN	141						
121	GEPOYDKSEVSHNLPRLTSRQ--DTSGEFSASPERLSVSTIAGKR--LPYSADVN	174						
142	GAPQAFQ-LNPVNPPLTNGOMDDIPREDHALVPSMG----GGKRIHLPY-ADDS	193						
175	QSPNRIVDP----VGLGNVAKKERVDGWMKKQEKNTGVPSTAASERGVYDIDASTD	228						
194	LPVOPRSDMPKSDLAAYGVSVMKEMEMWKKORER----NHQGNQGGDDGD--	245						
229	ILADEALLDNDARQPLSRKVSIPSSRIINPYRMYIMRLVTLCLFLYRTINPYNPAFAM	288						
246	--ADPLM-DEARQOLSRKILPSSQINPRMILITRLVGLGFFHYRMIHPANDAFAM	302						
289	LVSVCIEIWFSLWILQOFKPKFVPVNETLRLRLATRDREGEPSOLAADVITSVDTL	348						
303	LIVSICEIWFMSWILDQPKMFPIDRETYLDRLSLRFQEGQSQLARIDFVSTVDPL	362						
349	KEPPLVYANVYLSLAVDYPVYDKVSCYVPFDGAAMLSFESEIATSEPAKVPFCKYSI	408						
363	KEPPLVYTNVYLSLVDYPVYDKVSCVSDGGAAMTLFEALSTSEPAKVPFCKRYNI	422						
409	EPAPPEWFAKIDYLDKQVOTSEFVKORAMKREVEEFKIRIYALVSKALCKCPSEEGVMQ	468						
423	EPAPPEWFOOKIDYLDKVAANVRRRRAMKREYEEFKRIYALVAKKQVREBEVTMQ	482						
469	DGTPWPGNNTGDHPMTIQVPLGONGGLADAGNELPRLVLYSREKRPQFHHKKAGAMNL	528						
483	DGTPWPGNNVBDHPGMIOVELGSGGGLDCGNEPLRLVLYSREKRPQYHHKKAGAMNL	542						
529	VRSVAVLTNGEFLTLNDCDHYINNSKALREMGFLMDPNLGKQVCPVOPRPDGDIDKD	588						
543	VRSVAVLTNAYLTLNDCDHYINNSKALREMGCMMDPLGKKVVCYQFQRPQDGDIDKD	602						
589	RYANRNTVFEDINLRGLDGIQGPVYVGTGCVFNRTALYGERPIKVKHKKPRLSLKLCG-	647						
603	RYANRNTVFEDINNRGLDGIQGPVYVGTGCVFNRQALYGDAP--KTKRP--SPTCNC	657						
648	-----GSKKSKSAKKEBSDKK--SGRHTDSYVPEVPLDIDEGVEGACGDDEK	694						
658	WPKMCFCCCCGNGNRKKQTKTKPKTEKKKLLFFKKEEQSPAYVALGEIDEALPDA--ENRK	715						
695	ALLMSOMSLERFQSOAVFAVASTLMEGQVPSATPNLKLKAIYHYSQYEDKSPMGME	754						
716	AGIYNQOKLEKFEQSSVYVYVSTILLENGGTLKASAPSLKELAIYHYSQYEDKTMQME	775						
755	IGWYIGSVTEIDLTGFKMAHGWESIYCMKLPAPKGSAPINLSDRLNOVLRMAVGEI	814						
815	LESFHCPWVGYNRLKFLERFPAVUNVTYIPISIPILMYCTLLAACLFTNPQIIPQIN	874						
836	FFSNHCPWVGYNRLKFLERFPAVUNVTYIPISIPILMYCTLLAACLFTNPQIIPQIN	895						
875	IASIMFSLFLSIPATGILEMRSGVIGIDEMWRNEQFWYIGVSAHLFAVFOGILTVLNG	934						
896	VASIMFMSLFCIPATGILEMRSGVIGIDEMWRNEQFWYIGVSAHLFAVFOGILTVLNG	955						
935	IDTNFYVYTKRSDDDGPAFLYLFKWTLLIPPTLLIYNLVGVVAGVYATINSGVQSG	994						
956	VDTSFYVYTKSGD--DEESELTYFKWTLLIPPTLLIYNLVGVVAGVYATINSGVQSG	1014						

QY 995 PLFGKLEFAFWIYVHLVPLKGLMGKRONPTIIVVSVLLASIFSLMWRIDPPTSRYT 1054
 DB 1015 PLFGKLEFAFWIYVHLVPLKGLMGKRONPTIIVVSVLLASIFSLMWRIDPFLAKXD 1074
 QY 1055 GPDILECGINC 1065
 DB 1075 GPLLECGIDC 1085
 RESULT 12
 09LL12 PRELIMINARY; PRT; 1094 AA.
 AC 09LL12; 01-OCT-2000 (Tremblrel, 15, Created)
 DT 01-OCT-2000 (Tremblrel, 15, Last sequence update)
 DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
 DE Cellulose synthase-8.
 GN CESA-8.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Zea.
 OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=20398328; PubMed=10938350;
 RA Holland N., Holland D., Helentjaris T., Dhuga K.S.,
 RA Xocoostle-Cazares B., Delmer D.P.;
 RA "A comparative analysis of the plant cellulose synthase (cesa) gene
 RT family.";
 RL Plant Physiol. 123:1313-1324(2000).
 DR EMBL; AF200532; AAF89968.1; -
 DR InterPro; IPR001841; Znf_ring.
 DR InterPro; IPR001841; Znf_ring.
 DR Pfam; PF03552; Cellulose-synt; 1.
 DR SMART; SM00184; RING; 1.
 DR SMART; SM00184; RING; 1.
 SQ SEQUENCE 1094 AA; 122575 MW; DFBBC18F49D23F5E CRC64;
 Query Match 67.7%; Score 3843.5; DB 10; Length 1094;
 Best Local Similarity 66.3%; Pred. No. 16-289;
 Matches 727; Conservative 134; Mismatches 178; Indels 57; Gaps 17;
 QY 2 ESEGTACKPMKNIVPQCICSDNVGKTVGDRVACIGSPVCPCYCYERKDGNS 61
 DB 23 DRESGAAGGGAARRAARAPCOICGDEYGVGFDEPACNECAFPCVACYEREREGSOA 82
 QY 62 CPOCKTRVRLKGSAPIDKDEGLAD-----EGVEENVPQKEKISRMGLMHT 113
 DB 83 CPOCKTRVRLKGSAPIDKDEGLAD-----EGVEENVPQKEKISRMGLMHT 140
 QY 114 RKGGEEMGEPOYDKESHNLRLTSRQ---DTSGEFSASPERLSVSSTIAGKR--L 167
 DB 141 YGRGD-AHGFSPV---PNVPLLTNGCQVDDIPRQHLVSYNGGG--GGGRKIHPL 194
 QY 168 PVSVDVNSPNRRIYD-----VGLGNVAKERVDCGKMKOEKNTGPVSTQASERGV 221
 DB 195 PF-ADPNLNVQPSRMDPSKDLAAYGVSAVAKERMEGKOKER-----LQHVSESGG 247
 QY 222 DIDASTDLADALINDEAROPLSRKVSIPSSRINPVRVIMRLVITLFLHYRTINV 281
 DB 248 DMGDGD--ADPLM-DEAROPLSRKVSIPSSRINPVRVIMRLVITLFLHYRTINV 303
 QY 282 PNAFLMTVSVICETWIFALSWITLDOFKWFPVNRRETYDLRLALRYDEGEPSQLAAVD 341
 DB 304 KDAFLMTVSVICETWIFALSWITLDOFKWFPVNRRETYDLRLALRYDEGEPSQLAAVD 363
 QY 342 VSTVPLKEPPLVNTAVLSIAVDYPVDCVCIYFDDGAAMLSESTLAETSEFAKWP 401
 DB 364 VSTVPLKEPPLVNTAVLSIAVDYPVDCVCIYFDDGAAMLSESTLAETSEFAKWP 423
 QY 402 FCKKYSIEBRAEWEFAKIDYLDKQVQTSFYKDRAMKREVEERIRINALYSALCKP 461
 DB 424 FSKKNIEBRAEWEFAKIDYLDKQVQTSFYKDRAMKREVEERIRINALYSALCKP 483

QY 462 EEWVMDQGTPEWNGTGDHPGMIQVFLQNGGLDAEGNELPRLVYVSREKRSEFOHNRK 521
 DB 484 EEWVMDQGTPEWNGTGDHPGMIQVFLQNGGLDAEGNELPRLVYVSREKRSEFOHNRK 543
 QY 522 AGANALVAVSAVLNPGPILNDCHYINNSKALREACFLMDPMLGKOVCPORP 581
 DB 544 AGANALVAVSAVLNPGPILNDCHYINNSKALREACFLMDPMLGKOVCPORP 603
 QY 582 DGIDKNDRYANRVTVEFDINLNGLOIGPVPVYVGCVCNRRALGYEPPYKHKRPS 640
 DB 604 DGIDKNDRYANRVTVEFDINLNGLOIGPVPVYVGCVCNRRALGYEPPYKHKRPS 660
 QY 641 -----LTKLC--GGSRRKNSKAKESPSKSGRTHDTPVFNLDIIEGVEGAG 689
 DB 661 RHCNCPKMLCSCCCKRNNKKTTPKTEKKRRLFEKKAENDSPVYALGEIDEGAPG- 719
 QY 690 FDEKALMSQMSLEKRGQSAVFAVASTLMENGVPSATPENLKEAIVHISGCEYDKS 749
 DB 720 -DIEKAGIYNOCKLEKKEKFGQSSVFAVASTLMENGVPSATPENLKEAIVHISGCEYDKS 778
 QY 750 DMGMEIGWYGSVTEIDILGFMHARGMSIYCMPLPAFKGAPINLSDRLOVLRML 809
 DB 779 DMGMEIGWYGSVTEIDILGFMHARGMSIYCMPLPAFKGAPINLSDRLOVLRML 838
 QY 810 GSVEILFSRHCPIWYNGRLKFLERFAYVNTIYPTISIPILMYCTLLAVCLFTNOFTI 869
 DB 839 GSVEILFSRHCPIWYNGRLKFLERFAYVNTIYPTISIPILMYCTLLAVCLFTNOFTI 898
 QY 870 POISNATSWLSLPISTATGILEMRSGVIGIDEMRNEQVWIGVSAHLEFAYOGIL 929
 DB 899 PELTNVASTWLSLPISTATGILEMRSGVIGIDEMRNEQVWIGVSAHLEFAYOGIL 958
 QY 930 KYLAGIDNTFTYTSKASDDEGPAFLYLKMTLLIPPTLLIVNVGVVAGSYAING 989
 DB 959 KYLAGIDNTFTYTSKASDDEGPAFLYLKMTLLIPPTLLIVNVGVVAGSYAING 1017
 QY 990 YQSMGPLEFGKLEFAFWIYVHLVPLKGLMGKRONPTIIVVSVLLASIFSLMWRIDP 1049
 DB 1018 YQSMGPLEFGKLEFAFWIYVHLVPLKGLMGKRONPTIIVVSVLLASIFSLMWRIDP 1077
 QY 1050 TSVVTPGPDILECGINC 1065
 DB 1078 LAKSNGPLLECGIDC 1093
 RESULT 13
 09SWM6 PRELIMINARY; PRT; 1026 AA.
 AC 09SWM6; 01-MAY-2000 (Tremblrel, 13, Created)
 DT 01-MAY-2000 (Tremblrel, 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
 DE Cellulose synthase catalytic subunit.
 GN IRX3 OR TI0B6_80.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=CV. LANDSBERG ERRECTA;
 RA MEDLINE=99264300; PubMed=10330464;
 RA Taylor N.G., Scheible W.R., Cutler S., Somerville C.R., Turner S.R.;
 RA "The irregular xylem3 locus of Arabidopsis encodes a cellulose
 RT synthase required for secondary cell wall synthesis.";
 RL Plant Cell 11:769-780(1999).
 RP SEQUENCE FROM N.A.
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,
 RA Tabata S., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
 RA Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.

RM [3]
SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF091713; AAD40885.1;
DR EMBL: AL391142; CAC01737.1;
DR InterPro: IPR001510; Cellulose_synth.
DR InterPro: IPR001841; znf_ring.
DR Pfam: PF03552; Cellulose_synth; 1.
DR SMART: SM00184; Ring; 1.
SQ SEQUENCE 1026 AA; 115797 MW; 503BFBCT8BE6E511 CRC64;

Query Match 67.1%; Score 3811; DB 10; Length 1026;
Best Local Similarity 66.9%; Pred. No. 3.1e-287;
Matches 721; Conservative 115; Mismatches 143; Indels 98; Gaps 17;

QY 10 KPMKIVPOTQICSDNKGTVYDGRFVACDICSPPVCPCEYERKKGNSCPOCKTRY 69
DB 27 KPLKMLDQFCETICDQIGLVEGDLFVACNCGCPACRCPCEYERREGTONCPOCKTRY 86
QY 70 KRLKSPALPGDKDGLAD---EGTVEFNYPQKIKISERMIGMHLTGKGEKGE-POY 125
DB 87 KRLKSPVPEGEDEDEDIDIEYFNIEHEDDKHNSAEAMLYGKMSYGRGPEDENGRF 146
QY 126 DKEVSHNLPRLTSHQDTSGEFSASPERLSVSTIAG-----KRL-PPSSDVNQ 175
DB 147 PPVLAGH-----SGEF-----PVGGYGNGEGLHKRVHPYSSSENG 184
QY 176 SPNRITVPVGLGNVAMKERVDMKMKOEKNTGPVSTQAASRGVDIDASTDILADEAL 235
DB 185 SEG-----GMRERMDMKL-QHGNLGPPEP-----DDPEKGLI----- 216
QY 236 LNDEARPLSRKVSIPSSRINPYRMVIMRLVILCLFLHYRTNPVPAFALMVSICE 295
DB 217 --DEARQPLSRKVPASSKINRYAVIARLILAVFLRYRLNVDHALGMLSVICE 274
QY 296 IMFALSWILDOPPKMFVNRETYDLRLALRYDREGEPSQLAAVDIFVSTVDPKPPVLT 355
DB 275 IMFALSWILDOPPKMFIERETYLDRLLRYREGEPPMMLAVDVFVSTVDPKPPVLT 334
QY 356 ANTVLSIADVIVDKVSCYFYDDGAMLSFESELAETSEFARKVWPCKKYSIEPRADW 415
DB 335 SNTVLSIADMPYVPEKISCYVSDGASMLTFESLSETEAFARKVWPCKKFSIEPRADW 394
QY 416 YFAAKIDYLDKVOVSFYKDRBRAMKREYEEFKIRINALVSKALKCPREGVWQDSTPMPG 475
DB 395 YFTLVVDYLDQVHPTFVKERRAMKREYEEFKVIRINAVAKASKVPLEGWMQDSTPMPG 454
QY 476 NNTGDHPGMIOVFLGONGDLAEGNELPRLVYVSREKRPFGOHKKAGAMNALVVSAYL 535
DB 455 NNTKHPGMIOVFLGSHSGFDEVEGHELPRLVYVSREKRPFGOHKKAGAMNALVVSAYL 514
QY 536 TNGPPIILDCDHTYNNKALREAMCFIADPNLGAQVOCYVOEPQRFQDIDKNDRYANRT 595
DB 515 TNAPMLMLDCDHYVNNKAVREAMCFIADPNLGAQVOCYVOEPQRFQDIDTDRYANRT 574
QY 596 VFFDNLNLGDIQGVYVGTGCVENFNTALGYEPPVYKHKPSLLKSLG-----GSR 650
DB 575 VFFDNLNLGDIQGVYVGTGCVYKQALGYEPP--KGRKRPMTIS--GCCCCCFGR 630
QY 651 KNSKAKKESDKKSGRHTDSTVPYFNLDIDIEEGVEG--AGFDEKALMSQMSLEKRFQ 709
DB 631 RKNKK-----FSKNDNMNGVAAALGAGDEKDEHLMSENFKRTGQ 670
QY 710 SAVFASVTLMEGCVPPSPATPENLKEAIIHYISCTYEDKSDMGMEIGIYGSVTEIDIL 769
DB 671 SSIFATSTLMEGCGVPPSSPAVLLKEAIIHYISCTYEDKTEGTELGMIYGSITEDIL 730
QY 770 FMHARGRSITICMPKLPAPFGSAPINLSDRINOVLRMAISVELLFRRHCPIMWY-NG 828
DB 731 FMHARGRSITICMPKLPAPFGSAPINLSDRINOVLRMAISVELLFRRHCPIMWY-NG 790
QY 829 RLKFLERPAVYNTIYIPITSIPILMYCTILAVCLFTNQFIIPQISNINASIMWFLSLIF 888

DB 791 KKLWLERPAVYNTIYIPITSIPILMYCTILAVCLFTNQFIIPQISNINASIMWFLSLIF 850
QY 889 ATGILEMNSGCGIDEMWNRNEQFWYIGVSALHFPVGGILKVLGIDIONFNVTSAKSD 948
DB 851 VTGILELHNSGVSIEWMNRNEQFWYIGVSALHFPVGGILKVLGIDIONFNVTSAKSD 910
QY 949 DCDFAELVLFKWTTLIPPTTLIYNLVGVGVSAINSQSGMPLFGKLFPAFWYV 1008
DB 911 D-DFGELVAFKKTTLIPPTTLIYNLVGVGVSAINSQSGMPLFGKLFPAFWYV 969
QY 1009 HLYPRLKGLMGQRNPTPIVWVSVLASISFSLWVRIDPFTSRVTPGDIIECGINC 1065
DB 970 HLYPRLKGLMGQRNPTPIVWVSVLASISFSLWVRIDPFTSRVTPGDIIECGINC 1026

RESULT 14
Q9XHP6
ID Q9XHP6 PRELIMINARY; PRT; 1026 AA.
AC Q9XHP6;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Cellulose synthase catalytic subunit.
GN IRX3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=99. LANDSBERG, ERECTA.
RX MEDLINE=99264300; PubMed=10330464;
RA Taylor N.G., Schelble W.R., Cutler S., Somerville C.R., Turner S.R.;
RT "The irregular xylem3 locus of Arabidopsis encodes a cellulose
synthase required for secondary cell wall synthesis."
RL Plant Cell 11:769-780(1999).
DR EMBL: AF088917; AAD32031.1;
DR InterPro: IPR005150; Cellulose_synth.
DR InterPro: IPR001841; znf_ring.
DR Pfam: PF03552; Cellulose_synth; 1.
DR SMART: SM00184; Ring; 1.
SQ SEQUENCE 1026 AA; 115858 MW; 453BRDID283C4D70 CRC64;

Query Match 67.0%; Score 3805; DB 10; Length 1026;
Best Local Similarity 66.9%; Pred. No. 9e-287;
Matches 720; Conservative 115; Mismatches 144; Indels 98; Gaps 17;

QY 10 KPMKIVPOTQICSDNKGTVYDGRFVACDICSPPVCPCEYERKKGNSCPOCKTRY 69
DB 27 KPLKMLDQFCETICDQIGLVEGDLFVACNCGCPACRCPCEYERREGTONCPOCKTRY 86
QY 70 KRLKSPALPGDKDGLAD---EGTVEFNYPQKIKISERMIGMHLTGKGEKGE-POY 125
DB 87 KRLKSPVPEGEDEDEDIDIEYFNIEHEDDKHNSAEAMLYGKMSYGRGPEDENGRF 146
QY 126 DKEVSHNLPRLTSHQDTSGEFSASPERLSVSTIAG-----KRL-PPSSDVNQ 175
DB 147 PPVLAGH-----SGEF-----PVGGYGNGEGLHKRVHPYSSSENG 184
QY 176 SPNRITVPVGLGNVAMKERVDMKMKOEKNTGPVSTQAASRGVDIDASTDILADEAL 235
DB 185 SEG-----GMRERMDMKL-QHGNLGPPEP-----DDPEKGLI----- 216
QY 236 LNDEARPLSRKVSIPSSRINPYRMVIMRLVILCLFLHYRTNPVPAFALMVSICE 295
DB 217 --DEARQPLSRKVPASSKINRYAVIARLILAVFLRYRLNVDHALGMLSVICE 274
QY 296 IMFALSWILDOPPKMFVNRETYDLRLALRYDREGEPSQLAAVDIFVSTVDPKPPVLT 355
DB 275 IMFALSWILDOPPKMFIERETYLDRLLRYREGEPPMMLAVDVFVSTVDPKPPVLT 334


```

QY 356 ANTVLSILAVYFVDFKVCYVEDDGAAMLSFESLAESEFARKWVPCKKSYIEPRAPEM 415
DB 335 SNTVEIILAMDYVEKEISCVSDDGASMLTFESLSEAEPRKRVDFCKKFSIEPRAPEM 394
QY 416 YFAKIDYLDKDKVQTSFVDRBRAMKREYEFKIRIINALSKAKCEBEGVMDGTPMNG 475
DB 395 YFTLKVADYLDKDKVHPTFVKERRAMKREYEFKIRIINAQAKASKVPLEGIMDGTWPNG 454
QY 476 NNTGDHPRGMIOVFLGONGGLDAAGNELPRLVYVSREKRGFGOHKKAGANALVRSAYL 535
DB 455 NNTKDRPMIOVFLHSGSGFDVGHLPRLVYVSREKRGFGOHKKAGANALVRSAYL 514
QY 536 TNGPFLNLDCOHYTNNSKALREAPKCFGLADPNLCKQVYCFQFQPDGIDKNDRYANRNT 595
DB 515 TNAPFLNLDCOHYTNNSKAVREAPKCFGLADPNLCKQVYCFQFQPDGIDTNDRYANRNT 574
QY 596 VFEDILRGDGIQGVYVGTGCVFNRTALGYEPRITKVKHKRPSLSKLKG-----GSR 650
DB 575 VFEDIMKGLDGIQGVYVGTGCVFKRQALGYEPR--KGRPKRMIS--CGCCPQFGR 630
QY 651 KNSKAKKESDKKSGRHTDSTVPVNLDDIEGVEG--AGFDEKALMSOMSLERFGQ 709
DB 631 RKNKR-----FSKNDMNGDVALAGAECDKELHMFENMEKTFQG 670
QY 710 SAVFASITMENGVPSPATPENLKEAHLVHISCGEDKSDMCEMIGMIGYSTEDILNG 769
DB 671 SSIFVSTIMEBEGVSSPSPALLKEAHLVHISCGEDKTEMGTELGWIGYSTEDILNG 730
QY 770 FKMHARGMSIYCMPLPAFGKSPAINSLDRLNQVRMALGSEILFHSRCHPIWGY--NG 828
DB 731 FKMHGMSIYCMPLPAFGKSPAINSLDRLNQVRMALGSEILFHSRCHPIWGY--NG 790
QY 829 RLKFLERFAYVNTTITPITSIPMLAYCTLLAVCLFTNOPIPOISNIASIMPLSLPSIF 888
DB 791 KTKMLERFAYVNTTITPITSIPMLAYCTLLAVCLFTNOPIPOISNIASIMPLSLPSIF 850
QY 889 ATGILMRSGVIGIDEMMRNEQFVWIGVSAHLFANFQGLKLAGIDTNEFTYTSASDE 948
DB 851 VTGILMRSGVIGIDEMMRNEQFVWIGVSAHLFANFQGLKLAGIDTNEFTYTSASDE 910
QY 949 DCGFALYLFKMTTLIPPTLLIVLVYAGVSAVNSGYOSWGPFGKLFEPFANVIL 1008
DB 911 D-DEGLYAFKMTTLIPPTLLIVLVYAGVSAVNSGYOSWGPFGKLFEPFANVIL 969
QY 1009 HLYPFLKGLMGRNPTTIVVWSVLLASIFSLWVRIDPFTSRVTPDILEGGINC 1065
DB 970 HLYPFLKGLMGRNPTTIVVWSVLLASIFSLWVRIDPFTSRVTPDILEGGINC 1026

```

RESULT 15

09SKJ5 PRELIMINARY; PRT; 1065 AA.

```

ID 09SKJ5 AC 09SKJ5
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative cellulose synthase catalytic subunit.
GN ATG25540.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buehl C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Coppenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,

```

```

RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RP
RN [2]
RC SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006300; ADD20713.1; -
DR InterPro: IPR005150; Cellulose_synth.
DR InterPro: IPR001309; ICE_p20.
DR InterPro: IPR001841; ZnF_Cing.
DR Pfam: PF03552; Cellulose_synth; 1.
DR SMART: SM00184; RING; 1.
DR PROSITE: PS01121; CASPASE_HIS; UNKNOWN.1.
SQ
SEQUENCE 1065 AA; 120616 MW; F2AE36C9492A6076 CRC64;

Query Match 67.0%; Score 3801; DB 10; Length 1065;
Best Local Similarity 66.2%; Pred. No. 2e-286;
Matches 714; Conservative 132; Mismatches 156; Indels 76; Gaps 12;

QY 10 KPMKNVPTQICSDNNGKTYDGRFVACDICSFPVRCPYEERKDNOSQPOCKTAY 69
DB 23 KPLKDLNGOICQCGDDVLTGTGNVFAVACNCGFPLCSCYTEERKDSQCPCKANF 82
QY 70 KRLKSPALPGDKDEGLDDEGTVEFNYPQEKISERMGLNTRKGEEMGEPOYDKEY 129
DB 83 RRLNSPRLVDEKEDVDNDIEN-ERDYQ----- 111
QY 130 SHNHLPLRLTSRODTSGEFSAA--PERLSVSTIAGKRLPYSSDYNSPNR----- 179
DB 112 -GNKKARLPRAE---EFSSSRHRESLPVSLTHGH--PVSGEL-PPPDNATLSPCI 163
QY 180 -----RIVP-----VGLGVAMKRVGVMMKOKKNGPYSTOASBR 218
DB 164 DPQLPEITVOLLPLPRLIDPSKDLNLSYGVNDMKRIGKMLKODKNMHTGKHBEK 223
QY 219 GGVYDIDASTDILADEALNDEAROPLSRKSIPSRKINPYRMVIMLRVILDLFLHYRT 278
DB 224 GGE--PEGISNGDELQWVDARLPMSRVVHPSPAMPTRYRILVLRLLDVLFLHYRT 281
QY 279 NPYNAPFALMLVSYICEIFALSWIILDOPPKFPPVAKRETYLDRLALRYDREPSQLAAY 338
DB 282 HPKYKDAYALMLTSVICEIEMFASWMLDOPPKYPIKRETFLLRLALRYDRDEPSQLAAY 341
QY 339 DIFVSTVDDLPKEPPLTANTVLSILAVDPVQVSGVSGVDGGAAMLSEFSLFETSPARK 398
DB 342 DVFTSTVDPMKEPPLTANTVLSILAVDPVQVSGVSGVDGGAAMLSEFSLFETSPARK 401
QY 399 WPFCKKSYIEPRAPEMYPFAKIDYLDKDKVQTSFVDRBRAMKREYEFKIRIINALSKAL 458
DB 402 WPFCKKSYIEPRAPEMYPFAKIDYLDKDKVQTSFVDRBRAMKREYEFKIRIINALSKAL 461
QY 459 KCEBEGVMDGTPMFGNNTGHPGMIOVFLGONGGLDAAGNELPRLVYVSREKRGFGOH 518
DB 462 KIPEDGTEWDETSMPGNRDPHGMIOVFLGONGGLDAAGNELPRLVYVSREKRGFGOH 521
QY 519 HKKAGMANLVAVSAVLTNGPFLNLDCOHYTNNSKALREAPKCFGLADPNLCKQVYCFQ 578
DB 522 HKKAGMANLVAVSAVLTNGPFLNLDCOHYTNNSKALREAPKCFGLADPNLCKQVYCFQ 581
QY 579 QRFEDGIDKNDRYANRNTVFEDILRGDGIQGVYVGTGCVFNRTALGYEPRITKVKHKR 638
DB 582 QRFEDGIDKNDRYANRNTVFEDILRGDGIQGVYVGTGCVFNRTALGYEPRITKVKHKR 641
QY 639 PSLSKLKGLSGSRKNSKAF--KESDKKSGRHTDSTVPVNLDDIEGVEGAEFDEKALL 697
DB 642 PNIVYSCGSGSRKNSKAF--KESDKKSGRHTDSTVPVNLDDIEGVEGAEFDEKALL 699
QY 698 MSQMSLEKFGSGSAFVASTLMENGVPSPATPENLKEAHLVHISCGEDKSDMCEMIGM 757
DB 700 VSQKRLKREFGSPVFIATFMEQGLPSTNPLTLLEAHLVHISCGEDKSDMCEMIGM 759

```

```

QY 758 IYGSVTEEDILTGFMHARGMRSIYCMPLKPAFKGSAPINLSDRINOVLRNALGSVEILFS 817
    |||||
Db 760 IYGSVTEEDILTGFMHARGMRSIYCVSRPAFKGSAPINLSDRINOVLRNALGSVEILFS 819
    |||||
QY 818 RHCPIMWGYNGRLKFLERFAVNTTIVITSIPLLMYCTLAVCLFTNOFTIIPQISNIAS 877
    |||||
Db 820 RHCPIMWGYNGRLKFLERIAVINTIYVITSIPLLAYCMLPAFCLITNTFTIPEISNLAS 879
    |||||
QY 878 IWPISLFLSIFATGILEMRSGVGDIDEMWRNEQFWVIGVSAHLFAVFOGILKVLGIDT 937
    :|||
Db 880 LCFMLLFASISYASILELKWSDALEDWMRNEQFWVIGTSAHLEFAVFOGILKVFAGIDT 939
    :|||
QY 938 NFVTSKASDEDDGFAELYLKFTTLLIPTTLLIVNLGVVAGVSYAINSGYOSWGPLE 997
    |||||
Db 940 NFVTSKASDEDDGFAELYLKFTTLLIPTTLLIVNLGVVAGVSTAINSGYOSWGPLE 999
    |||||
QY 998 GKLEFAFWIVHLVPLKGLMGRONRPTIIVVWSVLLASISLWVRIDPFTSRVTG 1055
    |||||
Db 1000 GKLEFAFWIVHLVPLKGLMGRONRPTIIVVWSVLLASISLWVRIDPFTSRVTG 1056
    |||||

```

Search completed: June 16, 2003, 11:07:59
 Job time : 55 secs

THIS PAGE BLANK (USPTO)

T
..... (USPTO)